

THIS WEEK

EDITORIALS

WORLD VIEW Lacking motivation? Meet those who benefit **p.7**

QUIT Nicotine impact blocked in mice with injectable gene **p.8**

LIGHT SOURCE Invertebrates creep and crawl to street lamps **p.9**



Science takes the stand

Two legal rulings by the US Supreme Court last week will have significant implications for research into health-care outcomes and for how neuroscience is used in sentencing juveniles.

Many people were watching the US Supreme Court in Washington DC last week. The justices inside did not disappoint those who expected drama. Although the landmark (if rather complex) victory for President Barack Obama on health care took most headlines, a decision by the court a few days earlier also has significant implications. Neither judgment, it would seem at first glance, directly affects science and research. But both decisions deserve attention.

First, health care. With the court's decision to largely uphold the health-reform law, the nation moved much closer to providing health care for all — a provision that, beginning with nineteenth-century German health-care reformer Otto von Bismarck, almost all other wealthy nations have come to acknowledge as a right, not a privilege.

The decision also secures the future of several science-related provisions, assuming that congressional Republicans don't overturn the law legislatively. These include the launch of a non-profit, non-governmental Patient-Centered Outcomes Research Institute, based in Washington DC. Its remit is to compare the risks, benefits and effectiveness of different treatments and of various approaches to health-care delivery and management — a hugely important enterprise if US spending on health is to be brought under control.

The second decision could fundamentally change the criminal justice system for thousands of young people, but has altogether different implications for researchers. In deciding on two cases of crimes committed by people under 18 years old — *Miller v. Alabama* and *Jackson v. Hobbs* — the court voted 5–4 to prevent mandatory life sentences being given without the possibility of parole for juveniles convicted of murder. The decision rests on the fact that such sentences violate provisions in the Eighth Amendment to the Constitution that prohibit cruel and unusual punishment.

The opinion of the court notes that the decision was based not just on “common sense” about young people's immaturity but “on science and social science as well”. Although the research may simply be backing up conclusions that the court would have reached anyway, the language of the opinion suggests that the justices seriously considered scientific evidence as a factor.

Specifically, the judgment cites behavioural research and brain-imaging studies that show fundamental differences between the brains of adults and juveniles, the latter of which can be underdeveloped in areas that help to control impulses and avoid risk, among other behaviours. The court took the view that these factors could mitigate culpability, and suggest a higher chance of reform as young people mature. *Nature* reported on the difficulties of applying science in such cases in April (see *Nature* **484**, 304–306; 2012).

This is not the first time that scientific evidence has been used to bolster arguments for leniency towards young people: in a 2005 decision on *Roper v. Simmons*, the court took the death penalty off the table for young offenders, citing similar research. Now, the court says that it believes the behavioural and imaging evidence has got stronger.

The way the court fashioned its recent decision puts increasing pressure on researchers who study adolescent development to convert their research findings into a format that can be exploited to assess offenders. It does not prohibit life-without-parole sentences for young offenders

“An immature science could be increasingly drawn into the decision-making process.”

ers outright (as it had for the death penalty in *Roper v. Simmons*), but merely prevents states from making the punishment mandatory. This is laudable from a common-sense perspective because it means that punishments can be tailored to the specific crime and offender. But the ruling might mean that an immature science could be increasingly

drawn into the decision-making process. The psychological surveys and functional magnetic resonance imaging of brain structures cited in these cases are most relevant to population-level differences between juveniles and adults. Yet at the individual level, there is wide variation in how mature, culpable and capable of reform a particular offender is. There are no valid ways to predict where one person sits on these scales, yet the court's decision seems to endorse this approach.

Translating research from group findings to individuals is a challenge for many areas of society in which science helps to drive decisions: from medicine to environmental protection and in legal matters. Scientists in affected fields should consider this a call to arms. ■

Good advice

The UK government's latest appointment offers hope for British science.

Since its election two years ago, the UK government has sent out mixed signals on science and technology. Budgets for the grant-giving research councils have been maintained, but other areas of research spending have been cut. Meanwhile, the potential of science and innovation to help pull the economy out of recession features in government rhetoric from time to time, but not as frequently or emphatically as researchers had hoped. In this context, the announcement last week that Mark Walport, director of the Wellcome Trust, will next April assume the role of chief scientific adviser to the UK government, is being welcomed with an enthusiasm that goes far beyond the platitudes that usually greet such appointments.

Walport currently has one of the most powerful — not to mention best remunerated — positions in the world of science, responsible for the disbursement of more than £600 million (US\$940 million) annually at one of the world's largest research philanthropies. His readiness

to leave Wellcome's palatial London headquarters on the Euston Road for the uncertain pleasures of Whitehall strongly indicates that he has won satisfactory assurances from Prime Minister David Cameron and, perhaps, from chancellor George Osborne that they will actually listen to him. As such, the appointment itself seems to confound the widespread belief that the top echelons of Britain's Conservative–Liberal Democrat coalition government have minimal interest in science.

The chief scientific adviser's role in the UK government is a flexible one, not spelt out in legislation, and to a great extent the job is what the holder makes of it. The incumbent, population biologist John Beddington, has sought to strengthen networking between scientists and engineers inside government, and to encourage the appointment of scientific advisers in every government department. His public profile has been most strongly associated with two issues — climate change and food security — that were higher priorities for the previous, Labour government than they are for the coalition.

Walport's footprint can also be expected to reflect his own background. At Wellcome he has implemented a large and contentious shift away from small project and programme grants, and towards generous, long-term support for a few excellent researchers. He also had a key role in securing government backing for the planned Francis Crick Institute in central London (previously known as the UK Centre for Medical Research and Innovation).

These two efforts provide ample indication of what Walport can be expected to work for in government: heavier concentration of grant funding in the hands of the very best scientists and greater emphasis on 'translational' research. In both cases, Walport's perspectives seem to match those of David Willetts, the Conservative science minister.

It should be noted, of course, that the post of British chief scientific

adviser — unlike its US counterpart — is a non-political appointment. Walport will take up the position even in the unlikely, but possible, event of Cameron, Osborne and Willetts being voted out of power before next April.

He will arrive in the job some six months ahead of the next comprehensive spending review, which will determine the shape of British science in the medium term. The last such review, in 2010, allowed the Medical Research Council to grow its budget with inflation, by freezing (and so cutting, in real terms) other fields of science. There will be no easy options in 2013, but Walport's appointment will in itself raise hopes that some form of ring-fencing will continue to protect the overall science budget.

On the international scene, Walport will be expected to guide the government through some painful choices over global projects, such as Europe's Extremely Large Telescope, many of which Britain endorses but may be unable to pay to participate in.

As chief scientific adviser, Walport can also take a lead on issues within science itself, such as research misconduct and open-access publishing. In both areas, the research and university-funding councils have been criticized for failing to take any sort of stance. Walport can either tell them to do so or simply do it for them.

There are, of course, serious limitations on what one man, in a single office buried away in the Department for Business, Innovation and Skills, can do to change entrenched ways of thinking inside the British government — never mind in Britain itself. Still, in taking on the post, Walport lends much-needed credibility to the view that the chief scientific adviser might, indeed, make a difference. ■

"To a great extent, the chief scientific adviser's job is what the holder makes of it."

Print preview

The printing press changed the world; three-dimensional printing could do the same.

No science-fiction spaceship is complete without a replicator: a machine that, when fed with some nameless goo, can produce anything from a nutritious and delicious meal to a high-powered plasmatic continuum flux generator.

The real-world equivalent of the replicator is the three-dimensional (3D) printer, which can mass-produce replicas of everything from molecular structures to rare fossils. Not too long ago, 3D printing was extremely costly and time-consuming, and used a variety of exotic chemicals. These days, as the News Feature on page 22 reveals, it is a little less costly and takes less time, and can use a variety of 'inks', including silicone shower sealant.

It is hard even to guess the effects that 3D printing might have, not just on science, but also on manufacturing, construction, the economy and how we live our lives. Why go to a shop — or even online — to buy a gizmo, when you can print one at home? One can imagine the conflicts about intellectual property, similar to those that have changed the music industry beyond recognition and are now doing the same in publishing.

Still, if two-dimensional (2D) printing is anything to go by, the 3D version will suffer a number of teething problems before it gets to that stage. Engineers, after all, have yet to invent a cheap 2D printer that doesn't cost a fortune in ink cartridges or go wrong every five minutes.

Printing first came to Europe in the fifteenth century, when, as George Sampson said in *The Concise Cambridge History of English Literature* (Cambridge University Press, 1941), "upon the outworks of obstinate medievalism, rang out a series of hammer-strokes that shook the old world to pieces".

Johannes Gutenberg's first printed Bible appeared in Mainz, Germany, in 1455. "The coming of print is the most important event of the fifteenth century," said Sampson, because "as the pen is mightier than the sword, so the press is mightier than the pen". The earliest books in Europe were in Latin. But when William Caxton set up the first press in England, in 1476, he started to print books in English — often his own translations, with scholarly prefaces.

Soon, authors were queuing at Caxton's door to cast their own works into this dramatic medium. "After that I had accomplished and finished divers histories, as well of contemplation as of other historical and worldly acts of great conquerors and princes," wrote Caxton in the preface to his 1485 edition of Thomas Malory's *Le Morte D'Arthur*, "many noble and divers gentlemen of this realm of England came and demanded me many and oft times". These early printed books were, therefore, custom products, and their distribution was wide only relative to the hand-copied editions that had gone before them.

The rest, as they say, is history. Printing was perhaps the greatest driver of literacy there has ever been, and its effect on English was profound. Before printing, English was a collection of mutually almost unintelligible dialects, and those authors who used it wrote as they spoke. That modern readers can understand Geoffrey Chaucer's fourteenth-century *Canterbury Tales* without too much help is testament not to our cleverness, but to the fact that modern English grew out of the dialect spoken in London, where Chaucer wrote. *Sir Gawain and the Green Knight*, written by an anonymous contemporary of Chaucer in the north-west of England, would have puzzled Chaucer and is much harder work for us today. But thanks to printing, written English as we now know it had become more-or-less standardized by the seventeenth century.

Caxton's 2D printing set up was probably plagued with technical and mechanical problems, just as ours are. But the effect of printing on society, economics and language has been both profound and spectacular. Printing in three dimensions promises another such revolution, although in an entirely orthogonal direction. ■

➔ **NATURE.COM**
To comment online,
click on Editorials at:
go.nature.com/xbhunq

to leave Wellcome's palatial London headquarters on the Euston Road for the uncertain pleasures of Whitehall strongly indicates that he has won satisfactory assurances from Prime Minister David Cameron and, perhaps, from chancellor George Osborne that they will actually listen to him. As such, the appointment itself seems to confound the widespread belief that the top echelons of Britain's Conservative–Liberal Democrat coalition government have minimal interest in science.

The chief scientific adviser's role in the UK government is a flexible one, not spelt out in legislation, and to a great extent the job is what the holder makes of it. The incumbent, population biologist John Beddington, has sought to strengthen networking between scientists and engineers inside government, and to encourage the appointment of scientific advisers in every government department. His public profile has been most strongly associated with two issues — climate change and food security — that were higher priorities for the previous, Labour government than they are for the coalition.

Walport's footprint can also be expected to reflect his own background. At Wellcome he has implemented a large and contentious shift away from small project and programme grants, and towards generous, long-term support for a few excellent researchers. He also had a key role in securing government backing for the planned Francis Crick Institute in central London (previously known as the UK Centre for Medical Research and Innovation).

These two efforts provide ample indication of what Walport can be expected to work for in government: heavier concentration of grant funding in the hands of the very best scientists and greater emphasis on 'translational' research. In both cases, Walport's perspectives seem to match those of David Willetts, the Conservative science minister.

It should be noted, of course, that the post of British chief scientific

adviser — unlike its US counterpart — is a non-political appointment. Walport will take up the position even in the unlikely, but possible, event of Cameron, Osborne and Willetts being voted out of power before next April.

He will arrive in the job some six months ahead of the next comprehensive spending review, which will determine the shape of British science in the medium term. The last such review, in 2010, allowed the Medical Research Council to grow its budget with inflation, by freezing (and so cutting, in real terms) other fields of science. There will be no easy options in 2013, but Walport's appointment will in itself raise hopes that some form of ring-fencing will continue to protect the overall science budget.

On the international scene, Walport will be expected to guide the government through some painful choices over global projects, such as Europe's Extremely Large Telescope, many of which Britain endorses but may be unable to pay to participate in.

As chief scientific adviser, Walport can also take a lead on issues within science itself, such as research misconduct and open-access publishing. In both areas, the research and university-funding councils have been criticized for failing to take any sort of stance. Walport can either tell them to do so or simply do it for them.

There are, of course, serious limitations on what one man, in a single office buried away in the Department for Business, Innovation and Skills, can do to change entrenched ways of thinking inside the British government — never mind in Britain itself. Still, in taking on the post, Walport lends much-needed credibility to the view that the chief scientific adviser might, indeed, make a difference. ■

"To a great extent, the chief scientific adviser's job is what the holder makes of it."

Print preview

The printing press changed the world; three-dimensional printing could do the same.

No science-fiction spaceship is complete without a replicator: a machine that, when fed with some nameless goo, can produce anything from a nutritious and delicious meal to a high-powered plasmatic continuum flux generator.

The real-world equivalent of the replicator is the three-dimensional (3D) printer, which can mass-produce replicas of everything from molecular structures to rare fossils. Not too long ago, 3D printing was extremely costly and time-consuming, and used a variety of exotic chemicals. These days, as the News Feature on page 22 reveals, it is a little less costly and takes less time, and can use a variety of 'inks', including silicone shower sealant.

It is hard even to guess the effects that 3D printing might have, not just on science, but also on manufacturing, construction, the economy and how we live our lives. Why go to a shop — or even online — to buy a gizmo, when you can print one at home? One can imagine the conflicts about intellectual property, similar to those that have changed the music industry beyond recognition and are now doing the same in publishing.

Still, if two-dimensional (2D) printing is anything to go by, the 3D version will suffer a number of teething problems before it gets to that stage. Engineers, after all, have yet to invent a cheap 2D printer that doesn't cost a fortune in ink cartridges or go wrong every five minutes.

Printing first came to Europe in the fifteenth century, when, as George Sampson said in *The Concise Cambridge History of English Literature* (Cambridge University Press, 1941), "upon the outworks of obstinate medievalism, rang out a series of hammer-strokes that shook the old world to pieces".

Johannes Gutenberg's first printed Bible appeared in Mainz, Germany, in 1455. "The coming of print is the most important event of the fifteenth century," said Sampson, because "as the pen is mightier than the sword, so the press is mightier than the pen". The earliest books in Europe were in Latin. But when William Caxton set up the first press in England, in 1476, he started to print books in English — often his own translations, with scholarly prefaces.

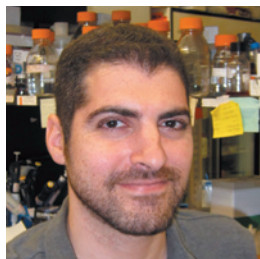
Soon, authors were queuing at Caxton's door to cast their own works into this dramatic medium. "After that I had accomplished and finished divers histories, as well of contemplation as of other historical and worldly acts of great conquerors and princes," wrote Caxton in the preface to his 1485 edition of Thomas Malory's *Le Morte D'Arthur*, "many noble and divers gentlemen of this realm of England came and demanded me many and oft times". These early printed books were, therefore, custom products, and their distribution was wide only relative to the hand-copied editions that had gone before them.

The rest, as they say, is history. Printing was perhaps the greatest driver of literacy there has ever been, and its effect on English was profound. Before printing, English was a collection of mutually almost unintelligible dialects, and those authors who used it wrote as they spoke. That modern readers can understand Geoffrey Chaucer's fourteenth-century *Canterbury Tales* without too much help is testament not to our cleverness, but to the fact that modern English grew out of the dialect spoken in London, where Chaucer wrote. *Sir Gawain and the Green Knight*, written by an anonymous contemporary of Chaucer in the north-west of England, would have puzzled Chaucer and is much harder work for us today. But thanks to printing, written English as we now know it had become more-or-less standardized by the seventeenth century.

Caxton's 2D printing set up was probably plagued with technical and mechanical problems, just as ours are. But the effect of printing on society, economics and language has been both profound and spectacular. Printing in three dimensions promises another such revolution, although in an entirely orthogonal direction. ■

➔ **NATURE.COM**
To comment online,
click on Editorials at:
go.nature.com/xbhunq

E. LUKYANOV



Meet patients to get your motivation back

Biomedical scientists risk forgetting what they're working for if they don't connect with the people who are affected by their research, says **Tal Nuriel**.

Most research scientists, especially in biomedicine, can probably remember when their early wide-eyed enthusiasm started to wane. For me, it was during my time as a research technician at New York University Langone Medical Center. One afternoon in the lab, I announced that I really wanted to “cure a disease” one day.

My comment was innocent and genuine, so I was caught off guard when a couple of postdocs in the lab laughed at it. They told me that researchers don't ‘cure’ anything any more; at best, they develop drugs effective enough to secure government approval, and even that is nearly impossible.

My idealism dampened further as I made my way through my graduate studies at the Weill Cornell Medical College in New York. My work centres around the effects of nitric oxide on Alzheimer's disease, and although my passion for the work is pure — my grandmother had the disease — I soon abandoned the idea that my research would ever make a real difference to actual people. The realities of research began to weigh on me, and with each failed experiment and exhausting late night, my focus in the lab shifted further away from trying to help people and closer towards trying to complete my thesis and publish enough papers to help me to secure a job after I graduated.

I found myself wondering whether there might be ways to halt this shift in focus and motivation, which so many young researchers experience. Then I found one. Back in January, in the middle of my seventh year of graduate studies (please shoot me), I enrolled in a ‘self-expression and leadership’ course, for which I had to organize a community project. And so it was that on a Thursday night in May, a little under 100 people directly affected by Alzheimer's gathered in a Manhattan bar to meet and mingle with about 20 young researchers, all of whom were trying to, if not cure the disease, then at least lessen its effects.

I called the event a Meet the Researchers night, and I was staggered by the interest. I invited patients and care-givers with the help of the local chapter of the Alzheimer's Association. Immediately after the invitations were sent out, we received a flood of responses, and ten further chapters said that they wanted to organize similar events in their cities. Apparently, meeting young scientists was more desirable than I had foreseen. And my colleagues were equally enthusiastic: with e-mails and help from administrative departments, I was able to recruit graduate students, postdocs and assistant professors from most of the major universities around New York.

At the event, all the scientists did an amazing job of welcoming the guests and being patient

with their questions. One sat with a rather intense older gentleman for nearly an hour, fielding question after question and smiling all the time. As most of the older guests left and the bar thinned out, many of the researchers turned their attention towards each other. At that point, I realized that an unexpected benefit of the event was that it brought together 20 young researchers from the same city and with the same research interests, to meet and interact for the first time.

For me, the most touching moment was when I met a woman called Beth, whose father had recently been diagnosed with Alzheimer's disease. We discussed research for a while — then Beth started to talk about her father and how his amazing memory had been a source of pride for him before the disease had struck. Her eyes began to fill with tears, and she stopped mid-sentence, the emotions pouring over her.

I put my hand on her shoulder and told her that it would be OK, comforting her as best I could. And as she regained her composure and continued with her stories about her dad, I remembered why I had organized this event in the first place. Although I had witnessed the effects of Alzheimer's disease first hand — both as a child, watching my grandmother's slow deterioration, and as an adult, witnessing my mother's anxiety about succumbing to the same fate — it had been a long time since I had directly felt the heartache of seeing a loved one slowly fade away. Talking to Beth, I reconnected with the pain and suffering that this disease causes, and with why research into it is so important. I remembered that Alzheimer's disease isn't just the abstract images of plaques and tangles that I show in my presentations. It is a real and devastating

condition and the only real hope for those affected by it — perhaps even for my own mother — is the research that my fellow scientists and I conduct.

In my opinion, this sort of event should be repeated on a much grander scale. It is crazy to me that, although the patient community spends countless hours raising money for researchers, and the research community spends countless hours working to find treatments for patients, there is actually almost no direct contact between these groups.

If you are a biomedical scientist, there are plenty of reasons to organize a Meet the Researchers event in your area. And I can tell you from experience that doing so will not only renew your motivation in the lab: it will also reconnect you with the reasons you got into science. ■

Tal Nuriel is a graduate student at the Weill Cornell Medical College in New York City.
e-mail: tan2003@med.cornell.edu

IT IS CRAZY TO ME
THAT THERE IS
ALMOST NO
**DIRECT
CONTACT**
BETWEEN PATIENTS
AND
RESEARCHERS.

➔ **NATURE.COM**
Discuss this article
online at:
go.nature.com/hh8qsk

SEVEN DAYS

The news in brief

BUSINESS

Pharma fines

Drug giant GlaxoSmithKline will pay US\$3 billion in fines for enormous health-care fraud. The firm, headquartered in London, is pleading guilty to promoting antidepressants for unapproved use, and to holding back data and making unsupported safety claims about its diabetes drug Avandia (rosiglitazone). The settlement was announced last November, but legal documents detailing the guilty pleas were unsealed on 2 July.

Private telescope

A non-profit foundation announced on 28 June that it plans to raise hundreds of millions of dollars to build a small space telescope that would orbit the Sun, discovering and tracking potentially hazardous Earth-crossing asteroids. The B612 Foundation, made up of engineers, planetary scientists, astronauts and former NASA officials, says that the telescope would be the world's first privately funded deep-space mission. See go.nature.com/7cspmy for more.

Obesity pill

The US Food and Drug Administration (FDA) approved a new weight-loss drug on 27 June, the first in

JOURNALISM AWARDS

Helen Pearson, *Nature's* chief features editor, last week won the Association of British Science Writers' best feature award for her article 'Study of a lifetime' (*Nature* **471**, 20–24; 2011). And in Britain's Online Media Awards (for all forms of journalism), *Nature's* website and Twitter feed both won commendations.



Pyre of ivory

Gabon burned its entire government stockpile of ivory — almost 5,000 kilograms of tusks and carvings — on 27 June, in a vivid display of its intent to crack down on illegal ivory trading. Last year, international recorded ivory seizures reached their highest point since a ban on

the ivory trade was established in 1989, and African elephant poaching levels are at their highest for a decade, according to a 21 June report from the United Nations' Convention on International Trade in Endangered Species of Wild Fauna and Flora.

more than a decade. Belviquin (lorcaserin), made by Arena Pharmaceuticals in San Diego, California, suppresses food cravings by mimicking the effects of serotonin in the brain. The agency rejected the drug two years ago because of safety concerns, but after additional tests it is now allowing Belviquin's use in obese adults and in overweight people who have a weight-related health condition. See go.nature.com/vajfqe for more.

Pharma closure

About 1,000 jobs are to be lost with the closure of pharmaceutical company Roche's research site in Nutley, New Jersey. Announcing the job losses on 26 June, Roche said the 80-year-old base would cease operation by the end of 2013; some of its research and development

will move to facilities in Switzerland and Germany, where around 80 jobs will be created. The firm, headquartered in Basel, Switzerland, also plans to build a translational-research centre on the east coast of the United States, creating about 240 jobs.

Amylin buy-up

Two pharmaceutical giants, Bristol-Myers Squibb and AstraZeneca, are teaming up to acquire the biotech firm Amylin, based in San Diego, California, for US\$5.3 billion (or \$7 billion with existing contractual obligations). Bristol-Myers Squibb, headquartered in New York City, is buying Amylin for its lucrative diabetes treatments. As part of the deal, announced on 29 June, Amylin will receive \$3.4 billion from London-based AstraZeneca, and the

two firms will share profits and losses from Amylin's drugs pipeline.

RESEARCH

Higgs hunt

Results from the Large Hadron Collider, due to be presented on 4 July at a conference on high-energy physics in Melbourne, Australia, will provide evidence of a new particle, physicists told *Nature* as this issue went to press. But, they added, more data would be needed to confirm whether the particle behaves exactly as the standard model of particle physics predicts for the long-sought Higgs boson. See go.nature.com/uzvtf for the latest news.

Cell bank to close

The Massachusetts Human Stem Cell Bank, at the University of Massachusetts

WILS YANICK MANIENGU/AF/GETTY IMAGES

G. WOOD/AFP/GETTY IMAGES

Medical School in Shrewsbury, will close when it runs out of public funding this year. The bank opened with state money in 2008, to allow work on newly derived human embryonic stem-cell lines while restrictions were in place on federal funding. But President Barack Obama's administration lifted the restrictions in 2009, obviating the need for the repository. See go.nature.com/beu3ww for more.

NIH funding

The US National Institutes of Health (NIH) will fund an initiative to help extramural researchers to diagnose mysterious maladies, and another to study how cells use RNA outside their membranes to communicate, the agency announced on 2 July. The two five-year projects, funded at US\$145 million and \$130 million, respectively, are the latest from the Common Fund, a pot of money worth \$545 million a year that the NIH director uses to make strategic investments. The two programmes will begin in 2013.

POLICY

Carbon tax

Australia introduced a carbon tax on 1 July, in which the country's 300 biggest emitters of carbon dioxide will pay



Aus\$23 (US\$23.5) per tonne emitted. The tax will increase by 2.5% a year above inflation until an emissions-trading scheme replaces it in 2015. But national newspaper polls reported considerable opposition; some 2,000 protesters marched against the tax in Sydney (pictured); and opposition politicians have vowed to scrap it if they win power, with elections due in late 2013.

Single EU patent

Politicians are still arguing over how to replace the current costly and fragmented European patent arrangements with a unified system, under which inventors can obtain a single patent for the entire region. Everything except the location of the patent court for ruling on disputes was agreed last year, and at a meeting in Brussels on 28–29 June, ministers from European Union member

states concluded that Paris, London and Munich would split the court. But Europe's parliament then postponed its vote on the law, complaining about the way it reduced the power of the European Court of Justice. See go.nature.com/yuiw8v for more.

Health act upheld

The US Supreme Court on 28 June voted narrowly in favour of upholding the Patient Protection and Affordable Care Act, including a key provision requiring everyone to buy health insurance. The law also includes provisions relevant to biomedical research, such as measures to speed up approval of generic versions of protein-based drugs. See page 13 for more.

UK science chief

Mark Walport will be the United Kingdom's next chief scientific adviser, the UK Cabinet Office announced on 28 June. Walport has been director of the Wellcome Trust, the United Kingdom's largest biomedical charity, since 2003. See page 20 for more.

Romania plagiarism

A Romanian academic committee has confirmed allegations of plagiarism in the PhD thesis of the country's prime minister, Victor Ponta. But the committee officially

COMING UP

6–10 JULY

In Ottawa, five of the world's largest societies for studying evolution and ecology team up for their first joint congress on evolutionary biology. go.nature.com/hwsqna

11–15 JULY

Europe's largest science conference, the Euroscience Open Forum, meets in Dublin. esof2012.org

ceased to exist before the end of its meeting on 29 June in Bucharest — it was disbanded by acting education minister Liviu Pop. That move has outraged Romanian academics. The government has referred the case to a national research-ethics council. See go.nature.com/nrdjp2 for more.

Nuclear restart

Japan restarted its first nuclear reactor, at the Ohi plant in Fukui prefecture, on 1 July — despite last-ditch protests against its reopening (see go.nature.com/tjylu4). Another Ohi reactor unit should also start this month. All of Japan's 48 other working reactors remain shuttered after the Fukushima nuclear disaster in March 2011.

Open access

The Wellcome Trust has made good on promises to enforce its open-access policy more rigorously. Since 2006, the trust has required research papers from work that it funds to be free to access within six months of publication. But the policy has only 55% compliance. On 28 June, the trust announced that those who fail to make papers open access will be refused funding renewals and new grants. See go.nature.com/5iizko for more.

NATURE.COM

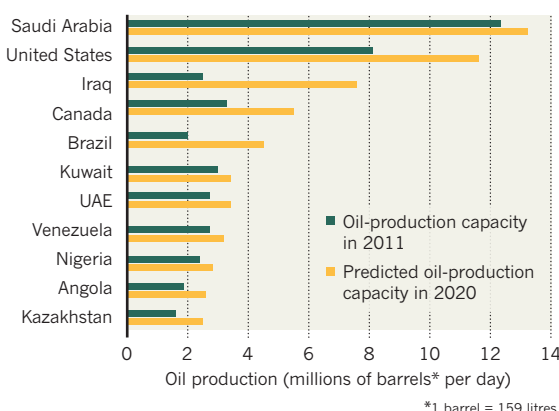
For daily news updates see:
www.nature.com/news

TREND WATCH

Several years of high prices are driving what could be the largest surge in global oil production since the 1980s, says a report from Harvard University's Belfer Center for Science and International Affairs in Cambridge, Massachusetts. Contrary to fears of a peak in oil production, the report suggests that global oil-production capacity could expand from 93 million barrels per day in 2011 to 110.6 million barrels per day in 2020. The increase is driven by advances in the production of 'unconventional' oil, such as that trapped in 'tight/shale' formations.

OIL-PRODUCTION SURGE

The production capacity of most oil-producing nations is set to increase dramatically by 2020, a study predicts, with the top 11 countries accounting for 80% of the potential new capacity.



SOURCE: BELFER CENTER

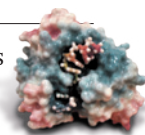
NEWS IN FOCUS

NUCLEAR ENERGY Proliferation fears greet project for cheap uranium enrichment **p.16**

AFRICA Dispute at flagship institute raises ghosts of colonial era **p.17**

MIDCONDUCT A data detective explains how he cracked his case **p.18**

TECHNOLOGY 3D printers give scientists a grip on research **p.22**



A. WONG/GETTY



Supporters of the Affordable Care Act celebrated last week, but the law cuts payments to US teaching hospitals, and could affect research.

HEALTH CARE

Health law worries hospitals

US academic centres fear they will lose out as upheld Affordable Care Act cuts payments.

BY MEREDITH WADMAN

A historic health-care decision by the US Supreme Court holds mixed blessings for research-intensive teaching hospitals. On 28 June, the court upheld most of the Patient Protection and Affordable Care Act, President Barack Obama's signature legislative accomplishment. By enabling tens of millions of currently uninsured people to obtain health-care coverage, the law should significantly lighten one of the hospitals' burdens: charity care. In 2010, 275 of the country's major teaching hospitals, which make up only 6% of all hospitals, conducted nearly 40% of the country's free care for uninsured patients, worth US\$8.4 billion, according to the Association

of American Medical Colleges (AAMC) in Washington DC.

Yet as broader insurance coverage reduces those costs, another of the law's provisions is ensuring that teaching hospitals do not receive a windfall. It cuts government health-care payments to hospitals by \$155 billion between 2010 and 2019 — reductions that teaching hospitals say may amount to a net loss, hurting their research mission. And although the cuts have already begun to take effect, the financial boost from an increasingly insured population won't begin until 2014, when people will be required to purchase health insurance or pay a tax penalty.

"There is no question that these cuts put the ability of institutions to continue to invest in

medical research at risk," says Ann Bonham, chief scientific officer of the AAMC.

Some health economists are sceptical of such claims. "All they are saying is: 'We want more, and we want other people to pay for it,'" says Michael Cannon, director of health-policy studies at the Cato Institute, a libertarian think tank in Washington DC. He believes that research should not be funded from the public purse. But Scott Gottlieb, a practicing physician and fellow at the conservative American Enterprise Institute, also in Washington DC, says that medical centres are right to be concerned. "They are not going to be able to make up in the volume of newly insured what they are losing in the cuts that they're facing," he says.

Gottlieb notes that the hospitals' situation ►

► may be exacerbated if too many states opt out of increasing the number of people covered under Medicaid, the government programme for the poor and disabled. The Supreme Court, in its only disagreement with the health-care act, threw out the section that would have allowed the federal government to financially punish states that reject the planned expansion — which was expected to cover at least 16 million people.

Where science is concerned, the court's narrow 5–4 decision upholding the law preserves several research efforts. They include the establishment of a \$10-billion Prevention and Public Health Fund, which the Obama administration is already using to prop up the budget of the Centers for Disease Control and Prevention in Atlanta, Georgia (see *Nature* **483**, 19; 2012), and which it is proposing to use to fund \$80 million in research into Alzheimer's disease at the National Institutes of Health (NIH). Other research provisions of the law include the Cures Acceleration Network, an NIH grants programme aimed at speeding into the clinic drugs and devices that industry has few incentives to develop, and the Patient-Centered Outcomes Research Institute in

Washington DC, which last month awarded its first \$30 million in grants for research comparing the effectiveness of different treatments. The law authorizes the Food and Drug Administration (FDA) to let makers of generic drugs compete with brand-name manufacturers in producing biosimilars — biological

“They are not going to be able to make up in the volume of newly insured what they are losing in the cuts that they’re facing.”

drugs based on large proteins (see *Nature* <http://doi.org/h2j>; 2012). That regulatory process is already well under way, and last month, Congress passed a bill establishing the first user fees to fund FDA approval of these drugs.

But researchers at teaching hospitals might be affected most by the \$155 billion in cuts to government payments from Medicaid and Medicare (which provides health insurance to people aged over 65). Already, biomedical research at teaching hospitals is partly supported by income from patient care, says Atul Grover, the AAMC's chief public-policy officer. If Medicare and Medicaid payments

are cut substantially, and new income from insured patients doesn't fill the gap, he says, research will suffer. “There's an old saying: no margin, no mission,” says Grover.

Edward Benz, president of the Dana-Farber Cancer Institute in Boston, Massachusetts, also questions whether increased revenues from newly insured patients will offset the cuts. He says that such patients will not necessarily continue to come to the teaching hospitals, because the law contains incentives that direct them to less costly hospitals and clinics.

Advocates of teaching hospitals point out that at least 20 million people will remain uninsured through choice or circumstance even after tens of millions of others obtain coverage under the law. A disproportionate number of the ill and injured among those without insurance may still end up at teaching hospitals, says Grover. Although the long wait for a ruling on the health-care law is over, uncertainties about the effects of the cuts to Medicare and Medicaid payments have just begun. Grover hopes that the cuts will be carefully adjusted in coming years to keep teaching hospitals on an even keel. “The devil is in the details here,” he says. ■

ENVIRONMENT

Palm-oil boom raises conservation concerns

Industry urged towards sustainable farming practices as rising demand drives deforestation.

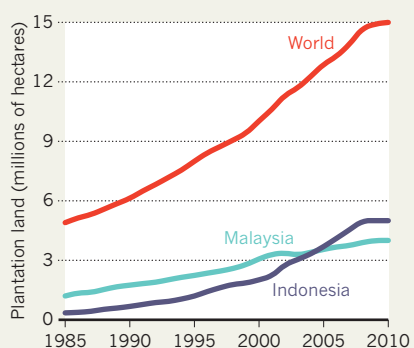
BY NATASHA GILBERT

Palm oil was once touted as a social and environmental panacea — a sustainable food crop, a biofuel that could help to cut greenhouse-gas emissions and a route out of poverty for small-scale farmers. In recent years, however, a growing body of research has questioned those credentials, presenting evidence that palm-oil farming can cause damaging deforestation and reduce biodiversity, and that the oil's use as a biofuel offers only marginal benefits for mitigating climate change.

But even as the environmental case against it grows stronger, the palm-oil business is booming as never before. “Oil palm is such a lucrative crop that there is almost no way to stop it,” says William Laurance, a forest-conservation scientist at James Cook University in Cairns, Australia. Indonesia, the world's largest grower of oil palms (see ‘Palm sprouts’), is expected to double production by 2030. And on 28 June, the Malaysian palm-oil company Felda Global

PALM SPROUTS

More than half of the world's palm-oil plants are farmed in Malaysia and Indonesia.



Ventures (FGV) earned US\$3.2 billion in the second-largest initial public offering (IPO) this year after Facebook, which will enable the company to bring thousands of extra hectares into production.

Sabri Ahmad, group president of FGV, told reporters last week that the company planned to expand its operations eightfold in eight years. To do so, it will have to look beyond Malaysia to countries such as Cambodia and Indonesia. Although Malaysia is now the world's second-largest producer of palm oil, it is running out of viable land for new oil-palm plantations, according to the US Department of Agriculture.

Such expansion is driven by the steadily rising demand for palm oil, mainly from the food sector, which uses it in a vast array of products, including margarine and biscuits. But the emerging biodiesel market is also thirsty for the oil.

In principle, biodiesel made from palm oil could be environmentally friendly, because the carbon dioxide released when it is burned is roughly the same as that absorbed as the plant grows. But vast swathes of forest have been cut down to make way for the crop, often in carbon-rich peatlands, where tree burning

SOURCE: FAO

► may be exacerbated if too many states opt out of increasing the number of people covered under Medicaid, the government programme for the poor and disabled. The Supreme Court, in its only disagreement with the health-care act, threw out the section that would have allowed the federal government to financially punish states that reject the planned expansion — which was expected to cover at least 16 million people.

Where science is concerned, the court's narrow 5–4 decision upholding the law preserves several research efforts. They include the establishment of a \$10-billion Prevention and Public Health Fund, which the Obama administration is already using to prop up the budget of the Centers for Disease Control and Prevention in Atlanta, Georgia (see *Nature* **483**, 19; 2012), and which it is proposing to use to fund \$80 million in research into Alzheimer's disease at the National Institutes of Health (NIH). Other research provisions of the law include the Cures Acceleration Network, an NIH grants programme aimed at speeding into the clinic drugs and devices that industry has few incentives to develop, and the Patient-Centered Outcomes Research Institute in

Washington DC, which last month awarded its first \$30 million in grants for research comparing the effectiveness of different treatments. The law authorizes the Food and Drug Administration (FDA) to let makers of generic drugs compete with brand-name manufacturers in producing biosimilars — biological

“They are not going to be able to make up in the volume of newly insured what they are losing in the cuts that they’re facing.”

drugs based on large proteins (see *Nature* <http://doi.org/h2j>; 2012). That regulatory process is already well under way, and last month, Congress passed a bill establishing the first user fees to fund FDA approval of these drugs.

But researchers at teaching hospitals might be affected most by the \$155 billion in cuts to government payments from Medicaid and Medicare (which provides health insurance to people aged over 65). Already, biomedical research at teaching hospitals is partly supported by income from patient care, says Atul Grover, the AAMC's chief public-policy officer. If Medicare and Medicaid payments

are cut substantially, and new income from insured patients doesn't fill the gap, he says, research will suffer. “There's an old saying: no margin, no mission,” says Grover.

Edward Benz, president of the Dana-Farber Cancer Institute in Boston, Massachusetts, also questions whether increased revenues from newly insured patients will offset the cuts. He says that such patients will not necessarily continue to come to the teaching hospitals, because the law contains incentives that direct them to less costly hospitals and clinics.

Advocates of teaching hospitals point out that at least 20 million people will remain uninsured through choice or circumstance even after tens of millions of others obtain coverage under the law. A disproportionate number of the ill and injured among those without insurance may still end up at teaching hospitals, says Grover. Although the long wait for a ruling on the health-care law is over, uncertainties about the effects of the cuts to Medicare and Medicaid payments have just begun. Grover hopes that the cuts will be carefully adjusted in coming years to keep teaching hospitals on an even keel. “The devil is in the details here,” he says. ■

ENVIRONMENT

Palm-oil boom raises conservation concerns

Industry urged towards sustainable farming practices as rising demand drives deforestation.

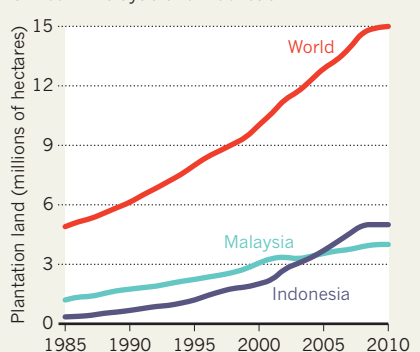
BY NATASHA GILBERT

Palm oil was once touted as a social and environmental panacea — a sustainable food crop, a biofuel that could help to cut greenhouse-gas emissions and a route out of poverty for small-scale farmers. In recent years, however, a growing body of research has questioned those credentials, presenting evidence that palm-oil farming can cause damaging deforestation and reduce biodiversity, and that the oil's use as a biofuel offers only marginal benefits for mitigating climate change.

But even as the environmental case against it grows stronger, the palm-oil business is booming as never before. “Oil palm is such a lucrative crop that there is almost no way to stop it,” says William Laurance, a forest-conservation scientist at James Cook University in Cairns, Australia. Indonesia, the world's largest grower of oil palms (see ‘Palm sprouts’), is expected to double production by 2030. And on 28 June, the Malaysian palm-oil company Felda Global

PALM SPROUTS

More than half of the world's palm-oil plants are farmed in Malaysia and Indonesia.



Ventures (FGV) earned US\$3.2 billion in the second-largest initial public offering (IPO) this year after Facebook, which will enable the company to bring thousands of extra hectares into production.

Sabri Ahmad, group president of FGV, told reporters last week that the company planned to expand its operations eightfold in eight years. To do so, it will have to look beyond Malaysia to countries such as Cambodia and Indonesia. Although Malaysia is now the world's second-largest producer of palm oil, it is running out of viable land for new oil-palm plantations, according to the US Department of Agriculture.

Such expansion is driven by the steadily rising demand for palm oil, mainly from the food sector, which uses it in a vast array of products, including margarine and biscuits. But the emerging biodiesel market is also thirsty for the oil.

In principle, biodiesel made from palm oil could be environmentally friendly, because the carbon dioxide released when it is burned is roughly the same as that absorbed as the plant grows. But vast swathes of forest have been cut down to make way for the crop, often in carbon-rich peatlands, where tree burning

SOURCE: FAO



Palm oil may offer a vital income to small-scale farmers, but its environmental impact is raising concerns.

and soil degradation release extra stores of the global-warming gas. A recent life-cycle assessment suggested that it could take up to 220 years for a plantation to become carbon neutral (W. M. J. Achten and L. V. Verchot *Ecol. Soc.* **16**, 14; 2011).

In January, after the US Environmental Protection Agency (EPA) found that palm-oil fuels emitted only 11–17% less greenhouse gas than diesel over their entire life cycle, it suggested that the oil should not be classified as a renewable fuel. Although a public consultation on the matter concluded in April, the EPA has not set a date to issue its final ruling. But the European Union (EU) continues to encourage the use of fuels based on palm oil. The EU has a binding target to raise the share of biofuels used in road transport to 10% by 2020, and most of that is expected to be met by blending biofuels such as palm oil with conventional fuels.

Research published in April also shows that oil-palm plantations are increasingly responsible for deforestation in Indonesia (K. M. Carlson *et al. Proc. Natl Acad. Sci. USA* **109**, 7559–7564; 2012). In Ketapang, an area in which the industry has been active, oil-palm

planting directly caused 27% of the region's deforestation in 2007–08. By 2020, around 40% of Ketapang will be given over to palm oil, up from 6% in 2007–08.

Palm oil would be much more sustainable if it were managed responsibly, says Nigel Sizer, director of the Global Forest Initiative at the World Resources Institute, an environmental think tank based in Washington DC. "It is possible to have carbon-neutral plantations if they are grown on already heavily logged and degraded land," he says.

Krystof Obidzinski, a forest-governance researcher at the Center for International Forestry Research in Bogor, Indonesia, agrees that there is plenty of non-forested or degraded land that could be used for plantations, but says that nations and companies need incentives to use it. Forested land is more attractive because companies can get extra income from the timber, and it is also less likely to be inhabited by large numbers of locals who can claim land rights and financial compensation, says Obidzinski.

Consumer pressure could encourage companies to change their practices. The Roundtable on Sustainable Palm Oil (RSPO), an

international non-profit association based in Zurich, Switzerland, that brings together conservation groups and palm-oil firms including FGV, says that it will not certify oil grown on land that was deforested to farm the crop. But many are sceptical that the RSPO, which was established in 2004, can effectively police the industry's rapid growth. "I am not confident at all that this is being done properly," says Joshua Linder, a biological anthropologist at the James Madison University in Harrisonburg, Virginia. He and others are keenly watching how the RSPO handles a flood of complaints filed this spring against a planned 70,000-hectare oil-palm plantation in Cameroon, for example, in what is seen as a test case of the body's power to hold companies to account.

In a statement, the RSPO told *Nature* that it was prepared to act in serious cases of environmental negligence, when it "may ultimately require a member to take specific actions or face cancellation of its RSPO membership".

But in the face of what Laurance calls "a green tidal wave" of oil-palm expansion, some fear that such sanctions will not be enough. ■



TOP STORY



Get the latest updates on the hunt for the Higgs with our News special go.nature.com/uzvtfid

MORE NEWS

- Genetically modified cotton gets high marks in India go.nature.com/uuhqe3
- Fossil hints that feathers may have been a feature of all dinosaurs go.nature.com/6zt7it
- Global network will track acidifying oceans go.nature.com/ywsxas

PODCAST



Spotting dark-matter filaments; and prenatal screening without dad go.nature.com/5efxhy

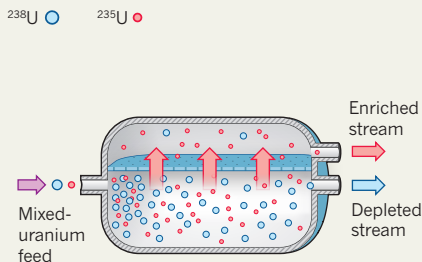
RISKY BUSINESS

Laser enrichment is cheaper and more efficient than other techniques for concentrating uranium-235 to make reactor fuel — but that could make it vulnerable to abuse, some non-proliferation experts fear.

GASEOUS DIFFUSION

Cost per SWU* **\$160** Energy cost Proliferation risk

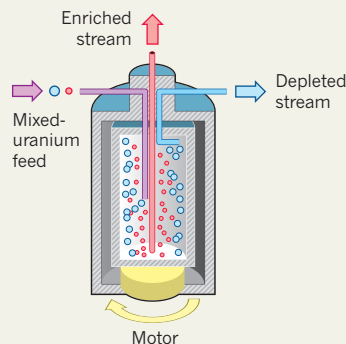
Because it is lighter and more active, ^{235}U is more likely than ^{238}U to bounce into walls and to cross a semi-permeable membrane.



GAS CENTRIFUGE

Cost per SWU* **\$100** Energy cost Proliferation risk

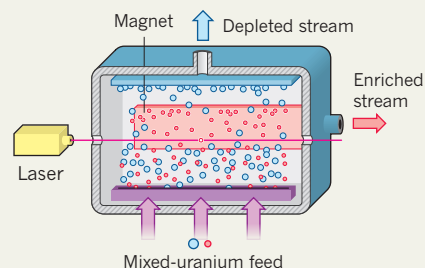
A spinning centrifuge pushes heavier ^{238}U towards the edges of the chamber, leaving a stream of enriched ^{235}U in the middle.



LASER ENRICHMENT

Cost per SWU* **\$30** Energy cost Proliferation risk

A tuneable laser excites and ionizes the ^{235}U in the mixed uranium feed. Magnets are then used to separate the ^{235}U from ^{238}U .



* Separative work unit

PROLIFERATION

Laser plant offers cheap way to make nuclear fuel

Experts worry that uranium-enrichment technique could be used to make bombs.

BY SHARON WEINBERGER

A controversial uranium-enrichment technology is on the cusp of making it cheaper to create fuel for nuclear power plants. But some non-proliferation experts are concerned that the efficiency of the laser-based technology will smooth the path for bomb-makers too.

On 11 July, the Atomic Safety and Licensing Board of the US Nuclear Regulatory Commission (NRC) will hold a final hearing on a proposal by General Electric (GE) of Fairfield, Connecticut, and Hitachi of Tokyo, Japan, to build the first commercial laser-enrichment plant. A decision on the plant, to be built in Wilmington, North Carolina, is anticipated in September and is widely expected to be favourable. But in a sign of the concerns that surround the technology, what is supposed to be the last public hurdle for the venture will be conducted in secret. “Although we would like to keep it as transparent as possible, the only practical thing to do with this mandatory review is to close this hearing in its entirety to the public,” says Paul Ryerson, one of the NRC’s administrative judges.

Separating the tiny fraction of uranium-235 from the uranium-238 that dominates natural uranium is the major hurdle to making fuel for commercial reactors and fissile material for weapons. Separation of Isotopes by Laser Excitation, or SILEX, a proprietary technique being developed by GE, promises to be much cheaper than either gaseous diffusion or gas centrifuging (see ‘Risky business’), two techniques currently in use that date back to the Manhattan Project. Although the exact details of SILEX are classified, the principles are well understood: a laser tuned to a specific frequency excites and ionizes the ^{235}U in a gaseous form of uranium, so that the charged atoms can be siphoned off.

GE and Hitachi are thought to be first companies to have sufficient skill with the process to build a commercial facility.

But many scientists and non-proliferation experts are concerned that a viable commercial facility would encourage countries wishing to start bomb projects. The American Physical Society in College Park, Maryland, for example, has lodged a

petition with the NRC, urging the commission to review proliferation risks for all licences. The society says that laser enrichment could be a “game changer” for those wanting to pursue proliferation, because it would fit into compact facilities — just one-quarter of the size of a centrifuge plant — and would therefore be difficult to detect through surveillance.

Scott Kemp, a nuclear expert at Princeton University in New Jersey, adds that many countries have a cadre of laser experts who could work on the technology. “That expertise does not exist for centrifuges, which are a bit esoteric,” he says.

CONFLICT OF INTERESTS

Former NRC commissioners disagree on whether the licensing process weighs up the proliferation risks carefully enough. Dale Klein, who was a commissioner during the early days of the GE–Hitachi proposal, says that proliferation risks were carefully considered for such applications. He notes, for example, that input was sought from the US defence department. But Victor Gilinsky, another former commissioner, says that other agencies involved in assessing proliferation

➔ **NATURE.COM**
For an opinion
article on nuclear
enrichment, see:
go.nature.com/vjztfh

risks were not regularly consulted. Gilinsky sees a tension between the United States' goal of safely commercializing nuclear-power technology and its efforts to control the proliferation of nuclear materials. "They are at cross purposes," he says. "When there's a conflict, generally speaking, the policy to spread nuclear technology overrides the non-proliferation policy."

GE spokesman Michael Tetuan says that the planned safeguards for the facility, such as measures to protect classified information, exceed the government's requirements. He also points to a report by an external panel, commissioned by GE but not made public,

which concluded that laser enrichment poses no greater proliferation risk than the other enrichment methods.

Donald Kerr, a former director of Los Alamos National Laboratory in New Mexico, who was a member of that panel, disagrees with the assessment that a laser facility would be smaller and more difficult to detect. The critics, he says, are relying on "marketing projects" from the 1980s that may have oversold the small footprint of the technology. "We had access to the actual information about the full-scale plant," says Kerr. The proposed plant would occupy 0.5 square kilometres. Kerr also dismisses concerns about industrial

espionage. "There's never been an American A. Q. Khan," he says, referring to the Pakistani nuclear scientist who stole industrial centrifuge secrets from the URENCO plant in the Netherlands, notoriously creating a nuclear black market.

Henry Sokolski, executive director of the Nonproliferation Policy Education Center in Washington DC, says that proliferation risks are harder to avoid than industrial espionage; the plant itself could simply spur other countries to follow suit and pursue their own research. "The most sensitive technology leak has already occurred," says Sokolski. "And it's that this stuff can work." ■ [SEE COMMENT P.30](#)

EMPLOYMENT

African researchers sue flagship programme for discrimination

Conflict at Kenya Medical Research Institute exposes widespread tensions.

BY LINDA NORDLING IN CAPE TOWN

The Kenya Medical Research Institute (KEMRI)–Wellcome Trust Research Programme is often seen as a model North–South partnership. African-run and mainly European-funded, it has trained dozens of African PhDs and done important research on malaria and other tropical diseases.

Yet in a court case that reawakens sour memories from colonial history, it now finds itself accused of exploiting African employees and holding back their careers compared with colleagues from developed countries. KEMRI denies the charges.

The case, which pits six African researchers, known as the KEMRI six, against the institute, highlights perceptions of unequal treatment that are common in joint programmes. Many prominent research institutions in Africa have evolved from field stations that once belonged to Europe, and although most are now owned locally, they remain dependent on funding and administrative support from their erstwhile masters.

"It is fair to say that this is an issue," says Marcel Tanner, director of the Swiss Tropical and Public Health Institute in Basel, which partners with health-research centres in Tanzania, Ivory Coast and Chad. Partnerships between rich and poor nations inevitably generate tensions, and those are exacerbated when the split between the haves and the have-nots runs along racial or former-colonial lines, says Kelly Chibale, a Zambia-born



Funding from Britain's Wellcome Trust is a cornerstone of the Kenya Medical Research Institute's programme in Kilifi.

biochemist who trained in Britain and the United States, and who now leads a drug-discovery centre at the University of Cape Town in South Africa. "There are tensions everywhere in science — but where the former colonial master is involved, it takes on a different dimension."

ALLEGATIONS OF BIAS

The KEMRI six — Samson Gwer, Michael Mwaniki, Nahashon Thuo, John Wagai, Moses Ndiritu and Albert Komba — were all medical officers or clinical research officers working towards PhDs, or about to start doing so, as part of the KEMRI–Wellcome Trust Research Programme in Kilifi, Kenya. The programme has a mixture of African and European staff, is run by the University of

Oxford, UK, and gets most of its funding from the Wellcome Trust, a British medical charity that has funded research in Africa for many decades. Neither body is a defendant in the case. "We are aware that allegations have been made concerning the KEMRI–Wellcome Trust research programme in Kenya," the trust said in a statement. "However, an investigation carried out by KEMRI found no evidence to support the serious allegations made by the researchers."

In their submission to Kenya's high court on 5 December 2011, the KEMRI six describe their treatment at the programme as "modern day slavery", alleging that they were passed over for promotion, training opportunities and grants while their white colleagues flourished. They also say that their work was stolen and given to researchers in the developed world, and that they were paid less than foreign colleagues with equivalent qualifications on the programme.

The six say that they raised their complaints with programme managers in 2010. Later that year they were suspended. They are suing KEMRI for compensation and demanding a court order for them to be reinstated unconditionally.

In submissions to the court, KEMRI says that the complaints from the six were dealt with in accordance with institutional employment policies. It also says that its staff are well paid in

➔ **NATURE.COM**
To read *Nature's*
Special on science in
Africa, see:
go.nature.com/ylnyfw

risks were not regularly consulted. Gilinsky sees a tension between the United States' goal of safely commercializing nuclear-power technology and its efforts to control the proliferation of nuclear materials. "They are at cross purposes," he says. "When there's a conflict, generally speaking, the policy to spread nuclear technology overrides the non-proliferation policy."

GE spokesman Michael Tetuan says that the planned safeguards for the facility, such as measures to protect classified information, exceed the government's requirements. He also points to a report by an external panel, commissioned by GE but not made public,

which concluded that laser enrichment poses no greater proliferation risk than the other enrichment methods.

Donald Kerr, a former director of Los Alamos National Laboratory in New Mexico, who was a member of that panel, disagrees with the assessment that a laser facility would be smaller and more difficult to detect. The critics, he says, are relying on "marketing projects" from the 1980s that may have oversold the small footprint of the technology. "We had access to the actual information about the full-scale plant," says Kerr. The proposed plant would occupy 0.5 square kilometres. Kerr also dismisses concerns about industrial

espionage. "There's never been an American A. Q. Khan," he says, referring to the Pakistani nuclear scientist who stole industrial centrifuge secrets from the URENCO plant in the Netherlands, notoriously creating a nuclear black market.

Henry Sokolski, executive director of the Nonproliferation Policy Education Center in Washington DC, says that proliferation risks are harder to avoid than industrial espionage; the plant itself could simply spur other countries to follow suit and pursue their own research. "The most sensitive technology leak has already occurred," says Sokolski. "And it's that this stuff can work." ■ [SEE COMMENT P.30](#)

EMPLOYMENT

African researchers sue flagship programme for discrimination

Conflict at Kenya Medical Research Institute exposes widespread tensions.

BY LINDA NORDLING IN CAPE TOWN

The Kenya Medical Research Institute (KEMRI)–Wellcome Trust Research Programme is often seen as a model North–South partnership. African-run and mainly European-funded, it has trained dozens of African PhDs and done important research on malaria and other tropical diseases.

Yet in a court case that reawakens sour memories from colonial history, it now finds itself accused of exploiting African employees and holding back their careers compared with colleagues from developed countries. KEMRI denies the charges.

The case, which pits six African researchers, known as the KEMRI six, against the institute, highlights perceptions of unequal treatment that are common in joint programmes. Many prominent research institutions in Africa have evolved from field stations that once belonged to Europe, and although most are now owned locally, they remain dependent on funding and administrative support from their erstwhile masters.

"It is fair to say that this is an issue," says Marcel Tanner, director of the Swiss Tropical and Public Health Institute in Basel, which partners with health-research centres in Tanzania, Ivory Coast and Chad. Partnerships between rich and poor nations inevitably generate tensions, and those are exacerbated when the split between the haves and the have-nots runs along racial or former-colonial lines, says Kelly Chibale, a Zambia-born



Funding from Britain's Wellcome Trust is a cornerstone of the Kenya Medical Research Institute's programme in Kilifi.

biochemist who trained in Britain and the United States, and who now leads a drug-discovery centre at the University of Cape Town in South Africa. "There are tensions everywhere in science — but where the former colonial master is involved, it takes on a different dimension."

ALLEGATIONS OF BIAS

The KEMRI six — Samson Gwer, Michael Mwaniki, Nahashon Thuo, John Wagai, Moses Ndiritu and Albert Komba — were all medical officers or clinical research officers working towards PhDs, or about to start doing so, as part of the KEMRI–Wellcome Trust Research Programme in Kilifi, Kenya. The programme has a mixture of African and European staff, is run by the University of

Oxford, UK, and gets most of its funding from the Wellcome Trust, a British medical charity that has funded research in Africa for many decades. Neither body is a defendant in the case. "We are aware that allegations have been made concerning the KEMRI–Wellcome Trust research programme in Kenya," the trust said in a statement. "However, an investigation carried out by KEMRI found no evidence to support the serious allegations made by the researchers."

In their submission to Kenya's high court on 5 December 2011, the KEMRI six describe their treatment at the programme as "modern day slavery", alleging that they were passed over for promotion, training opportunities and grants while their white colleagues flourished. They also say that their work was stolen and given to researchers in the developed world, and that they were paid less than foreign colleagues with equivalent qualifications on the programme.

The six say that they raised their complaints with programme managers in 2010. Later that year they were suspended. They are suing KEMRI for compensation and demanding a court order for them to be reinstated unconditionally.

In submissions to the court, KEMRI says that the complaints from the six were dealt with in accordance with institutional employment policies. It also says that its staff are well paid in

➔ **NATURE.COM**
To read *Nature's*
Special on science in
Africa, see:
go.nature.com/ylnyfw

► Kenyan terms, and that allegations of the theft of intellectual-property rights are unfair and unsubstantiated. Last week, the court asked both parties to provide more evidence; the next hearing is scheduled for 20 September.

SUPPORT NETWORK

Other African researchers at KEMRI have rallied to the programme's defence. The programme has sponsored more than 30 Africans for master's degrees in the past five years, and a further 35 research assistants have won external master's fellowships totalling around £2 million (US\$3.1 million), mainly from the Wellcome Trust. An £8-million strategic award from the Wellcome Trust is boosting PhD training on the programme, which currently has almost 50 PhD students. The investment puts it ahead of any other equivalent programme in the country, says Abdisalan Noor, who leads the programme's spatial-epidemiology group. "This has saved many Kenyan students from the vagaries of chasing postgraduate funding through the limited international and national scholarship opportunities."

In its submission to the court, KEMRI lists a number of African scientists who were promoted and developed by the programme. These include Charles Mbogo, deputy director of the Centre of Geographical Medicine Research

"There are tensions everywhere in science — but where the former colonial master is involved, it takes on a different dimension."

Coast in Kilifi, which houses the KEMRI–Wellcome Trust programme; and Gilbert Kokwaro, who went on to head the programme's pharmacology group and is now director of the Consortium for National Health Research, a non-profit organization that coordinates health research and training in Kenya. In 2008, the KEMRI–Wellcome programme hired Kenyan malaria researcher Samson Kinyanjui as head of training to improve support for African scientists.

However, outsiders have questioned whether the programme is doing enough to promote Africans. In 2010, senior international scientists independently reviewed the programme's

application for core funding for 2011–16 from the Wellcome Trust, and unanimously deemed it "excellent". But they also queried why the programme seemed to be so "separate" from the rest of KEMRI, and why so few senior African scientists were involved.

Some joint programmes have found it challenging to boost the number of African scientists. Not only are African researchers relatively scarce, it is often difficult for those who have trained in Africa to compete on standard metrics — publication and citation rates — with colleagues trained and well-connected in developed countries.

But a culture of openness and trust may help to avoid conflict in partnerships between African and developed countries, says Tanner, who helped to develop guidelines published in May by the Commission for Research Partnerships with Developing Countries in Berne, Switzerland. One approach, he says, is to let all partners participate in setting the research agenda.

"It's about respect and transparency," says Tanner. "If you have that, you can build up a good partnership." ■

Q&A Uri Simonsohn

The data detective

Psychology was already under scrutiny following a series of high-profile controversies. Now it faces fresh questions over research practices that can sometimes produce eye-catching — but irreproducible — results. Last week, Erasmus University Rotterdam in the Netherlands said that social psychologist Dirk Smeesters had resigned after an investigation found that he had massaged data to produce positive outcomes in his research, such as the effect of colour on consumer behaviour^{1,2}. Smeesters says the practices he used are common in the field. None of his co-authors is implicated. The university was tipped off by social psychologist Uri Simonsohn at the University of Pennsylvania in Philadelphia, who spoke exclusively to Nature about his investigation.

How did your investigation begin, and how did you analyse the papers?

Somebody sent me a paper by Smeesters. I was working on another project on false positives and had become pretty good at picking up on the tricks that people pull to get a positive result³. With the Smeesters paper, I couldn't find any red flags, but there were really far-fetched predictions.

The basic idea is to see if the data are too close to the theoretical prediction, or if multiple estimates are too similar to each other. I looked at several papers by Smeesters and asked him for the raw data, which he sent. I did some additional analyses on those and the results looked less likely. I'll be submitting a paper on the method this week.

I shared my analyses with Smeesters, showing him that the data didn't look real, and I offered

several times to explain my methods. He said he was going to re-run the study and retract the paper. That was all I heard until December, when Erasmus University Rotterdam contacted me and asked me to tell them why I was suspicious. They had started their own investigation.

Can we expect more cases like this?

I tried my approach with Diederik Stapel's data after he had been called out for fraud (see *Nature* 479, 15; 2011), and they looked fake from the very beginning. Besides him and Smeesters, there's another person. I found three suspicious papers, engaged him for several months, and eventually contacted the university. They had already started an investigation, which



LUKE CHURCH PHOTOGRAPHY

has ended. It's not official yet.

There's a fourth case in which I am convinced that there's fabrication. I've approached co-authors, but none of them wanted to help. If I didn't have anything else to do, I'd do something about it, but it just became too difficult because I was handling these other cases and my own research. It's very draining.

Is this indicative of deeper problems in the field?

I don't know how systemic the crime is. What's systemic is the lack of defences. Social psychology — and science in general — doesn't have sufficient mechanisms for preventing fraud. I doubt that fabrication is any worse in psychology than in other fields. But I'm worried by how easy it was for me to come across these people.

► **NATURE.COM**
Psychology's replication problems:
go.nature.com/mn4hdk

► Kenyan terms, and that allegations of the theft of intellectual-property rights are unfair and unsubstantiated. Last week, the court asked both parties to provide more evidence; the next hearing is scheduled for 20 September.

SUPPORT NETWORK

Other African researchers at KEMRI have rallied to the programme's defence. The programme has sponsored more than 30 Africans for master's degrees in the past five years, and a further 35 research assistants have won external master's fellowships totalling around £2 million (US\$3.1 million), mainly from the Wellcome Trust. An £8-million strategic award from the Wellcome Trust is boosting PhD training on the programme, which currently has almost 50 PhD students. The investment puts it ahead of any other equivalent programme in the country, says Abdisalan Noor, who leads the programme's spatial-epidemiology group. "This has saved many Kenyan students from the vagaries of chasing postgraduate funding through the limited international and national scholarship opportunities."

In its submission to the court, KEMRI lists a number of African scientists who were promoted and developed by the programme. These include Charles Mbogo, deputy director of the Centre of Geographical Medicine Research

"There are tensions everywhere in science — but where the former colonial master is involved, it takes on a different dimension."

Coast in Kilifi, which houses the KEMRI–Wellcome Trust programme; and Gilbert Kokwaro, who went on to head the programme's pharmacology group and is now director of the Consortium for National Health Research, a non-profit organization that coordinates health research and training in Kenya. In 2008, the KEMRI–Wellcome programme hired Kenyan malaria researcher Samson Kinyanjui as head of training to improve support for African scientists.

However, outsiders have questioned whether the programme is doing enough to promote Africans. In 2010, senior international scientists independently reviewed the programme's

application for core funding for 2011–16 from the Wellcome Trust, and unanimously deemed it "excellent". But they also queried why the programme seemed to be so "separate" from the rest of KEMRI, and why so few senior African scientists were involved.

Some joint programmes have found it challenging to boost the number of African scientists. Not only are African researchers relatively scarce, it is often difficult for those who have trained in Africa to compete on standard metrics — publication and citation rates — with colleagues trained and well-connected in developed countries.

But a culture of openness and trust may help to avoid conflict in partnerships between African and developed countries, says Tanner, who helped to develop guidelines published in May by the Commission for Research Partnerships with Developing Countries in Berne, Switzerland. One approach, he says, is to let all partners participate in setting the research agenda.

"It's about respect and transparency," says Tanner. "If you have that, you can build up a good partnership." ■

Q&A Uri Simonsohn

The data detective

Psychology was already under scrutiny following a series of high-profile controversies. Now it faces fresh questions over research practices that can sometimes produce eye-catching — but irreproducible — results. Last week, Erasmus University Rotterdam in the Netherlands said that social psychologist Dirk Smeesters had resigned after an investigation found that he had massaged data to produce positive outcomes in his research, such as the effect of colour on consumer behaviour^{1,2}. Smeesters says the practices he used are common in the field. None of his co-authors is implicated. The university was tipped off by social psychologist Uri Simonsohn at the University of Pennsylvania in Philadelphia, who spoke exclusively to Nature about his investigation.

How did your investigation begin, and how did you analyse the papers?

Somebody sent me a paper by Smeesters. I was working on another project on false positives and had become pretty good at picking up on the tricks that people pull to get a positive result³. With the Smeesters paper, I couldn't find any red flags, but there were really far-fetched predictions.

The basic idea is to see if the data are too close to the theoretical prediction, or if multiple estimates are too similar to each other. I looked at several papers by Smeesters and asked him for the raw data, which he sent. I did some additional analyses on those and the results looked less likely. I'll be submitting a paper on the method this week.

I shared my analyses with Smeesters, showing him that the data didn't look real, and I offered

several times to explain my methods. He said he was going to re-run the study and retract the paper. That was all I heard until December, when Erasmus University Rotterdam contacted me and asked me to tell them why I was suspicious. They had started their own investigation.

Can we expect more cases like this?

I tried my approach with Diederik Stapel's data after he had been called out for fraud (see *Nature* 479, 15; 2011), and they looked fake from the very beginning. Besides him and Smeesters, there's another person. I found three suspicious papers, engaged him for several months, and eventually contacted the university. They had already started an investigation, which



LUKE CHURCH PHOTOGRAPHY

has ended. It's not official yet.

There's a fourth case in which I am convinced that there's fabrication. I've approached co-authors, but none of them wanted to help. If I didn't have anything else to do, I'd do something about it, but it just became too difficult because I was handling these other cases and my own research. It's very draining.

Is this indicative of deeper problems in the field?

I don't know how systemic the crime is. What's systemic is the lack of defences. Social psychology — and science in general — doesn't have sufficient mechanisms for preventing fraud. I doubt that fabrication is any worse in psychology than in other fields. But I'm worried by how easy it was for me to come across these people.

Do you worry about other psychologists' reactions to your investigations?

I did worry a lot. Everybody likes the fact that whistle-blowers exist, but nobody likes them. People worry about somebody engaging in a witch-hunt, but I have a technique that is accurate, I used it when confronted with evidence, and I subjected it to replication by checking other papers from the same author. That's no more a witch-hunt than a neighbour calling the police when someone breaks into another person's home. I did not take justice into my own hands, I contacted the authorities and they took care of the rest. I suspect some people will be against what I've done, but there is really no personal benefit to someone of doing what I am doing.

So what is your motivation?

Simply that it is wrong to look the other way. If there's a tool to detect fake data, I'd like people to know about it so we can take findings that aren't true out of our journals. And if it becomes clear that fabrication is not an unusual event, it will be easier for journals to require authors to publish all their raw data. It's extremely hard for fabrication to go undetected if people can look at your data.

A university's reputation suffers a lot when people fake data, but they don't have tools for preventing that — journals do. Journals should be embarrassed when they publish fake data, but there's no stigma. They're portrayed as the victims, but they're more like the facilitators, like a government that looks the other way. I'd like journals to take ownership of the problem and start working towards stopping it.

Previous challenges to data in psychology were made by internal whistle-blowers, but you are not connected to Smeesters. Does that herald an important change?

It's a very important difference. The tool should be broadly applicable to other disciplines. I think it'll be worthwhile to find other ways of finding fake data. We know people are really bad at emulating random data, so there should be all sorts of tests that could be developed.

Is it possible that such methods could falsely ensnare innocent researchers?

That's my biggest fear; it's why I look at different papers from the same person. I wouldn't contact anybody unless they had three suspicious papers. And before any concerns become public, a proper investigation should always take place. ■

INTERVIEW BY ED YONG

1. Johnson, C. S., Smeesters, D. & Wheeler, S. C. *J. Pers. Soc. Psychol.* **102**, 32–50 (2012).
2. Smeesters, D. & Liu, J. J. *Exp. Soc. Psychol.* **47**, 653–656 (2011).
3. Simmons, J. P., Nelson, L. D. & Simonsohn, U. *Psychol. Sci.* **22**, 1359–1366 (2011).



More than 700 people protested at the trial of political scientist Büşra Ersanlı in Istanbul this week.

HUMAN RIGHTS

Turkey cracks down on academic freedom

External groups hope scrutiny will restrain government.

BY ALISON ABBOTT

Turkey is upping the pressure on scientists and students who question its policies, and international human-rights advocates are taking notice.

In the past few years, the government has clamped down on the independence of the Scientific and Technological Research Council of Turkey and the Turkish Academy of Sciences (see *Nature* **477**, 131; 2011). It has also harassed and jailed individual academics and students. Now, an international network is launching a campaign to support Turkish scientists whose academic rights it considers to have been violated. The network has issued a report and this week carried out its first concerted street action, when more than 100 of its supporters joined a large protest at the opening of the trial of Büşra Ersanlı, a political scientist at Marmara University in Istanbul.

Ersanlı was arrested last October, under

Turkey's 2006 anti-terrorist laws. A member of the legal Peace and Democracy Party, which promotes the rights of Turkey's Kurdish minority, she denies charges of supporting an outlawed separatist terrorist organization, the Kurdish Workers' Party.

Authorities have tried to prevent other scientists from speaking out against industrial interests, says Nesrin Uçarlar, a political scientist who has worked with Ersanlı at Marmara University. One targeted researcher is Onur Hamzaoglu, an epidemiologist at Kocaeli University in İzmit, who revealed that the region's industrial basin has high pollution levels and increased cancer rates. Hamzaoglu is now being investigated for unethical behaviour leading to public alarm, and faces a jail sentence.

Ersanlı will be tried alongside 204 others charged with illegally promoting Kurdish rights. Her arrest prompted colleagues in France to launch the International Workgroup on Academic Liberty and Freedom of Research in Turkey (GIT) on 21 November. The group is also drawing attention to the more than ▶

► **NATURE.COM**
For an Editorial on human rights in Turkey, see:
go.nature.com/v5jhez

Do you worry about other psychologists' reactions to your investigations?

I did worry a lot. Everybody likes the fact that whistle-blowers exist, but nobody likes them. People worry about somebody engaging in a witch-hunt, but I have a technique that is accurate, I used it when confronted with evidence, and I subjected it to replication by checking other papers from the same author. That's no more a witch-hunt than a neighbour calling the police when someone breaks into another person's home. I did not take justice into my own hands, I contacted the authorities and they took care of the rest. I suspect some people will be against what I've done, but there is really no personal benefit to someone of doing what I am doing.

So what is your motivation?

Simply that it is wrong to look the other way. If there's a tool to detect fake data, I'd like people to know about it so we can take findings that aren't true out of our journals. And if it becomes clear that fabrication is not an unusual event, it will be easier for journals to require authors to publish all their raw data. It's extremely hard for fabrication to go undetected if people can look at your data.

A university's reputation suffers a lot when people fake data, but they don't have tools for preventing that — journals do. Journals should be embarrassed when they publish fake data, but there's no stigma. They're portrayed as the victims, but they're more like the facilitators, like a government that looks the other way. I'd like journals to take ownership of the problem and start working towards stopping it.

Previous challenges to data in psychology were made by internal whistle-blowers, but you are not connected to Smeesters. Does that herald an important change?

It's a very important difference. The tool should be broadly applicable to other disciplines. I think it'll be worthwhile to find other ways of finding fake data. We know people are really bad at emulating random data, so there should be all sorts of tests that could be developed.

Is it possible that such methods could falsely ensnare innocent researchers?

That's my biggest fear; it's why I look at different papers from the same person. I wouldn't contact anybody unless they had three suspicious papers. And before any concerns become public, a proper investigation should always take place. ■

INTERVIEW BY ED YONG

1. Johnson, C. S., Smeesters, D. & Wheeler, S. C. *J. Pers. Soc. Psychol.* **102**, 32–50 (2012).
2. Smeesters, D. & Liu, J. J. *Exp. Soc. Psychol.* **47**, 653–656 (2011).
3. Simmons, J. P., Nelson, L. D. & Simonsohn, U. *Psychol. Sci.* **22**, 1359–1366 (2011).



More than 700 people protested at the trial of political scientist Büşra Ersanlı in Istanbul this week.

HUMAN RIGHTS

Turkey cracks down on academic freedom

External groups hope scrutiny will restrain government.

BY ALISON ABBOTT

Turkey is upping the pressure on scientists and students who question its policies, and international human-rights advocates are taking notice.

In the past few years, the government has clamped down on the independence of the Scientific and Technological Research Council of Turkey and the Turkish Academy of Sciences (see *Nature* **477**, 131; 2011). It has also harassed and jailed individual academics and students. Now, an international network is launching a campaign to support Turkish scientists whose academic rights it considers to have been violated. The network has issued a report and this week carried out its first concerted street action, when more than 100 of its supporters joined a large protest at the opening of the trial of Büşra Ersanlı, a political scientist at Marmara University in Istanbul.

Ersanlı was arrested last October, under

Turkey's 2006 anti-terrorist laws. A member of the legal Peace and Democracy Party, which promotes the rights of Turkey's Kurdish minority, she denies charges of supporting an outlawed separatist terrorist organization, the Kurdish Workers' Party.

Authorities have tried to prevent other scientists from speaking out against industrial interests, says Nesrin Uçarlar, a political scientist who has worked with Ersanlı at Marmara University. One targeted researcher is Onur Hamzaoglu, an epidemiologist at Kocaeli University in İzmit, who revealed that the region's industrial basin has high pollution levels and increased cancer rates. Hamzaoglu is now being investigated for unethical behaviour leading to public alarm, and faces a jail sentence.

Ersanlı will be tried alongside 204 others charged with illegally promoting Kurdish rights. Her arrest prompted colleagues in France to launch the International Workgroup on Academic Liberty and Freedom of Research in Turkey (GIT) on 21 November. The group is also drawing attention to the more than ▶

► **NATURE.COM**
For an Editorial on human rights in Turkey, see:
go.nature.com/v5jhez

► 770 students who are in prison in Turkey, most arrested for protesting against government policies, including the introduction of university fees.

"Freedom of expression in academia is being increasingly clamped down on in Turkey," says Vincent Duclert, a historian at the School for Advanced Studies in the Social Sciences in Paris and a founder of the network. "The GIT initiative will keep the cases in the eye of the international public, and this may bring pressure on Turkish authorities."

Uçarlar, whose PhD thesis on Kurdish language rights was derecognized by Marmara University in 2008, launched GIT's node in Turkey last December. Since then, six further branches have been established in North America and Europe.

GIT Turkey issued a report last week listing a selection of cases of academic rights violations, from arrests on terrorist charges to dismissals for trade-union membership and prosecution for statements deemed harmful to industry. "We don't really know how many cases are out there, because university staff are afraid to speak out," says Uçarlar. "But now we have a platform, people are starting to contact us about their experiences." The group has sent its report to the European Parliament, and has asked that the European Union's annual report on Turkey's bid for membership rates the country on freedom of expression for academics.

Erol Gelenbe, a computer scientist at Imperial College London who was educated in Turkey, points out that although the erosion of academic freedom in the country has accelerated in the past two years, "there has always been little tolerance for independent thinking". He says that at different times over the past few decades, "academics have been expelled from universities either because they were to the right or because they were to the left of the particular government".

Uçarlar agrees that the political situation is "complicated", with right-wingers, left-wingers and staunch secularists all under attack. "I never agreed with the policies of Kemal Gürüz," she says, referring to a former president of the Turkish higher-education council who enforced a ban on wearing the Islamic headscarf in Turkish universities. "But I'm appalled he was arrested on 25 June without credible charges."

Ayşe Erzan, a GIT supporter and a physicist at the Istanbul Technical University who co-founded an alternative science academy when the Turkish Academy of Sciences was put under government management last year, says that "people are getting nervous and there is a feeling we need to resist further slide into undemocratic measures". ■

POLICY

Britain names next chief science adviser

Immunologist Mark Walport, head of one of the world's largest biomedical charities, will take on role in 2013.

BY EWEN CALLAWAY

Biomedical research may soon enjoy an even higher profile in Britain's science-policy agenda. Mark Walport, director of the Wellcome Trust in London, will become the next government chief scientific adviser (GCSA), it was announced last week. He will replace population biologist John Beddington from April 2013.

Walport, a 59-year-old immunologist who took the helm at Wellcome in 2003, will be the first biomedical scientist in 30 years to occupy the role, providing advice directly to Prime Minister David Cameron and the Cabinet. As the outgoing head of one of the world's wealthiest biomedical-research charities, commanding an annual budget of more than £600 million (US\$940 million), he will be one of the most influential scientists to hold the position.

"I think it's going to be really important in terms of assuring that biomedical science is very high on the agenda in the government," says Nic Jones, chief scientist at the London-based charity Cancer Research UK. Although the GCSA does not oversee the research councils that dole out most government science funding, Jones hopes that Walport will make a strong case for biomedical and translational research in the government's next spending review, expected in the next two years.

The UK coalition government has already shown favour for biomedicine over other sciences. It largely protected the Medical Research Council's £800-million research budget in the latest spending review, and a December 2011 life-sciences strategy — which Walport helped to develop — emphasizes translational-research initiatives such as drug repurposing and stem-cell therapies. But Walport will need to be an advocate for science policy-making throughout government, says James Wilsdon, a science-policy expert at the University of Sussex in Brighton, UK. "The key thing will be that he moves above and beyond issues and areas where he's already been associated, to act as a broader leader across the community, across a broader set of issues."

Wilsdon and others expect that, just as Beddington championed government policies to address global food security and climate change, Walport will push his own favoured issues, such

as open access to research literature and data. Last week, the Wellcome Trust announced that it may withhold funding to grant recipients who do not make their research available within six months of publication — a policy instituted under Walport's watch. Open access is already gaining traction in the UK government (see *Nature* 486, 302–303; 2012), and with Walport beating the same drum, "that creates a real opportunity for more rapid and far-reaching progress to embed these sorts of changes in the research community," says Wilsdon.

One of the key tests for a GCSA is how they deal with crises outside their areas of expertise, says Colin Blakemore, a neuroscientist at the University of Oxford, UK, and former chief executive of the Medical Research Council. Beddington, for example, was confronted with

challenges including a global H1N1 influenza pandemic and a volcanic ash plume that clogged up Britain's airspace. Walport, says Blakemore, has both the breadth of knowledge and the access to expertise to confront such challenges.

At the Wellcome Trust, he showed his mettle by pushing through a controversial change to funding policy, targeting more cash to outstanding scientists rather than to individual research projects. He also boosted the trust's commitment to public engagement, the arts and the humanities.

Walport's appointment has unleashed speculation about who will replace him as head of the Wellcome Trust. Insiders say that the hunt for a successor will consider scientists with backgrounds in either academia or industry, and will be explicitly international. "I would expect a global search, a big-hitting shortlist," says Wilsdon. "It is now one of the biggest jobs in international science." ■ [SEE EDITORIAL P.5](#)

"The key thing will be that he moves above and beyond issues and areas where he's already been associated."

CORRECTION

The News story 'Fetal tests spur legal battle' (*Nature* 486, 454; 2012) located medical-genomics company Natera in Redwood City, California. The firm has actually relocated to San Carlos, California.

► 770 students who are in prison in Turkey, most arrested for protesting against government policies, including the introduction of university fees.

"Freedom of expression in academia is being increasingly clamped down on in Turkey," says Vincent Duclert, a historian at the School for Advanced Studies in the Social Sciences in Paris and a founder of the network. "The GIT initiative will keep the cases in the eye of the international public, and this may bring pressure on Turkish authorities."

Uçarlar, whose PhD thesis on Kurdish language rights was derecognized by Marmara University in 2008, launched GIT's node in Turkey last December. Since then, six further branches have been established in North America and Europe.

GIT Turkey issued a report last week listing a selection of cases of academic rights violations, from arrests on terrorist charges to dismissals for trade-union membership and prosecution for statements deemed harmful to industry. "We don't really know how many cases are out there, because university staff are afraid to speak out," says Uçarlar. "But now we have a platform, people are starting to contact us about their experiences." The group has sent its report to the European Parliament, and has asked that the European Union's annual report on Turkey's bid for membership rates the country on freedom of expression for academics.

Erol Gelenbe, a computer scientist at Imperial College London who was educated in Turkey, points out that although the erosion of academic freedom in the country has accelerated in the past two years, "there has always been little tolerance for independent thinking". He says that at different times over the past few decades, "academics have been expelled from universities either because they were to the right or because they were to the left of the particular government".

Uçarlar agrees that the political situation is "complicated", with right-wingers, left-wingers and staunch secularists all under attack. "I never agreed with the policies of Kemal Gürüz," she says, referring to a former president of the Turkish higher-education council who enforced a ban on wearing the Islamic headscarf in Turkish universities. "But I'm appalled he was arrested on 25 June without credible charges."

Ayşe Erzan, a GIT supporter and a physicist at the Istanbul Technical University who co-founded an alternative science academy when the Turkish Academy of Sciences was put under government management last year, says that "people are getting nervous and there is a feeling we need to resist further slide into undemocratic measures". ■

POLICY

Britain names next chief science adviser

Immunologist Mark Walport, head of one of the world's largest biomedical charities, will take on role in 2013.

BY EWEN CALLAWAY

Biomedical research may soon enjoy an even higher profile in Britain's science-policy agenda. Mark Walport, director of the Wellcome Trust in London, will become the next government chief scientific adviser (GCSA), it was announced last week. He will replace population biologist John Beddington from April 2013.

Walport, a 59-year-old immunologist who took the helm at Wellcome in 2003, will be the first biomedical scientist in 30 years to occupy the role, providing advice directly to Prime Minister David Cameron and the Cabinet. As the outgoing head of one of the world's wealthiest biomedical-research charities, commanding an annual budget of more than £600 million (US\$940 million), he will be one of the most influential scientists to hold the position.

"I think it's going to be really important in terms of assuring that biomedical science is very high on the agenda in the government," says Nic Jones, chief scientist at the London-based charity Cancer Research UK. Although the GCSA does not oversee the research councils that dole out most government science funding, Jones hopes that Walport will make a strong case for biomedical and translational research in the government's next spending review, expected in the next two years.

The UK coalition government has already shown favour for biomedicine over other sciences. It largely protected the Medical Research Council's £800-million research budget in the latest spending review, and a December 2011 life-sciences strategy — which Walport helped to develop — emphasizes translational-research initiatives such as drug repurposing and stem-cell therapies. But Walport will need to be an advocate for science policy-making throughout government, says James Wilsdon, a science-policy expert at the University of Sussex in Brighton, UK. "The key thing will be that he moves above and beyond issues and areas where he's already been associated, to act as a broader leader across the community, across a broader set of issues."

Wilsdon and others expect that, just as Beddington championed government policies to address global food security and climate change, Walport will push his own favoured issues, such

as open access to research literature and data. Last week, the Wellcome Trust announced that it may withhold funding to grant recipients who do not make their research available within six months of publication — a policy instituted under Walport's watch. Open access is already gaining traction in the UK government (see *Nature* 486, 302–303; 2012), and with Walport beating the same drum, "that creates a real opportunity for more rapid and far-reaching progress to embed these sorts of changes in the research community," says Wilsdon.

One of the key tests for a GCSA is how they deal with crises outside their areas of expertise, says Colin Blakemore, a neuroscientist at the University of Oxford, UK, and former chief executive of the Medical Research Council. Beddington, for example, was confronted with

challenges including a global H1N1 influenza pandemic and a volcanic ash plume that clogged up Britain's airspace. Walport, says Blakemore, has both the breadth of knowledge and the access to expertise to confront such challenges.

At the Wellcome Trust, he showed his mettle by pushing through a controversial change to funding policy, targeting more cash to outstanding scientists rather than to individual research projects. He also boosted the trust's commitment to public engagement, the arts and the humanities.

Walport's appointment has unleashed speculation about who will replace him as head of the Wellcome Trust. Insiders say that the hunt for a successor will consider scientists with backgrounds in either academia or industry, and will be explicitly international. "I would expect a global search, a big-hitting shortlist," says Wilsdon. "It is now one of the biggest jobs in international science." ■ **SEE EDITORIAL P.5**

"The key thing will be that he moves above and beyond issues and areas where he's already been associated."

CORRECTION

The News story 'Fetal tests spur legal battle' (*Nature* 486, 454; 2012) located medical-genomics company Natera in Redwood City, California. The firm has actually relocated to San Carlos, California.



BY NICOLA JONES

THREE-DIMENSIONAL PRINTERS ARE OPENING UP NEW WORLDS TO RESEARCH.

Christoph Zollikofer witnessed the first birth of a Neanderthal in the modern age. In his anthropology lab at the University of Zurich, Switzerland, in 2007, the skull of a baby *Homo neanderthalensis* emerged from a photocopier-sized machine after a 20-hour noisy but painless delivery of whirring motors and spitting plastic. This modern miracle had endured a lengthy gestation: it took years for Zollikofer's collaborators to find suitable bones from a Neanderthal neonate, analyse them with a computed-tomography (CT) scanner and digitally stitch them together on the computer screen. The labour, however, was simple: Zollikofer just pressed 'print' on his lab's US\$50,000 three-dimensional (3D) printer.

A pioneer in the use of 3D printing for research, Zollikofer started 20 years ago with a prototype that was even more expensive and required toxic materials and solvents — limitations that put off most scientists. But now newer, cheaper technology is catching on. Just as an inkjet printer sprays ink onto a page line by line, many modern 3D devices spray material — usually plastic — layer by layer onto a surface, building up a shape. Others fuse solid layers out of a vat of liquid or powdered plastic, often using ultraviolet or infrared light. Any complex shape can be printed, sometimes with the help of temporary scaffolding that is later dissolved or chipped away. These days, personal kits go for as little as \$500, says Terry Wohlers, a consultant and market analyst based in Fort Collins, Colorado — although industrial systems cost an average of \$73,000. Last year, he says, nearly 30,000 printers were sold worldwide, with academic institutions buying one-third of those in the \$15,000–30,000 price range.

Early adopters are using the technology to investigate complex molecules, fashion custom lab tools, share rare artefacts and even print cardiac tissue that beats like a heart. At palaeontology and anthropology conferences, more and more people are carrying printouts of their favourite fossils or bones. "Anyone who thinks of themselves as an anthropologist

needs the right computer graphics and a 3D printer. Otherwise it's like being a geneticist without a sequencer," says Zollikofer.

The printouts are yielding insights that are not possible with more conventional methods. Neanderthal neonate fossils, for example, are extremely rare, so Zollikofer did not want to risk copying his fragile specimen with the usual plaster-casting methods. With the printout, however, Zollikofer could explore the logistics of Neanderthal births. Along with the neonate skull, he printed out an adult, female Neanderthal pelvis and literally re-enacted a delivery. Some researchers had speculated that Neanderthals' wide hips made labour easier than it is for modern humans, but Zollikofer's experiment showed that the bigger skulls of Neanderthal neonates counteracted that advantage (M. S. Ponce de León *et al. Proc. Natl Acad. Sci. USA* **105**, 13764–13768; 2008). Like humans today, Neanderthals had the biggest heads — and brains — possible at birth, giving them a jump-start on development.

In his work, Zollikofer swaps back and forth between printed models and virtual ones. The computer models are good for calculating volumes or piecing together bone fragments — researchers can position them in space without gravity causing them to fall. But with the virtual models, he says, "you lose the sensation of touch, and even a notion of the size of the fossils". The physical models are far better for seeing how pieces should fit together in the first place, he adds.

MOLECULAR PLAYGROUND

Chemists and molecular biologists have long used models to get a feel for molecular structures and make sense of X-ray and crystallography data. Just look at James Watson and Francis Crick, who in 1953 made their seminal discovery of DNA's structure with the help of a rickety construction of balls and sticks.

Printed models help to reveal how molecules function.

These days, 3D printing is being used to mock up far more complex systems, says Arthur Olson, who founded the molecular graphics lab at the Scripps Research Institute in La Jolla, California, 30 years ago. These include molecular environments made up of thousands of interacting proteins, which would be onerous-to-impossible to make any other way. With 3D printers, Olson says, “anybody can make a custom model”. But not everybody does: many researchers lack easy access to a printer, aren’t aware of the option or can’t afford the printouts (which can cost \$100 or more).

Yet Olson says that these models can bring important insights. When he printed out one protein for a colleague, they found a curvy ‘tunnel’ of empty space running right through it. The conduit couldn’t be seen clearly on the computer screen, but a puff of air blown into one side of the model emerged from the other. Determining the length of such tunnels can help researchers to work out whether, and how, those channels transport molecules. Doing that on the computer would have required some new code; with a model, a bit of string did the trick.

Software that lets researchers twist and turn such structures on a computer screen is extremely useful, says Olson, but inadequate. Even the most advanced software will let two atoms occupy the same space. And tinkering with molecules inside a computer is a grind — it takes time for the computer to re-render an object after every turn, and interpreting the pictures requires mental effort. Fiddling with a physical model, on the other hand, is more like play. “I don’t have to think about it; I just do it,” says Olson.

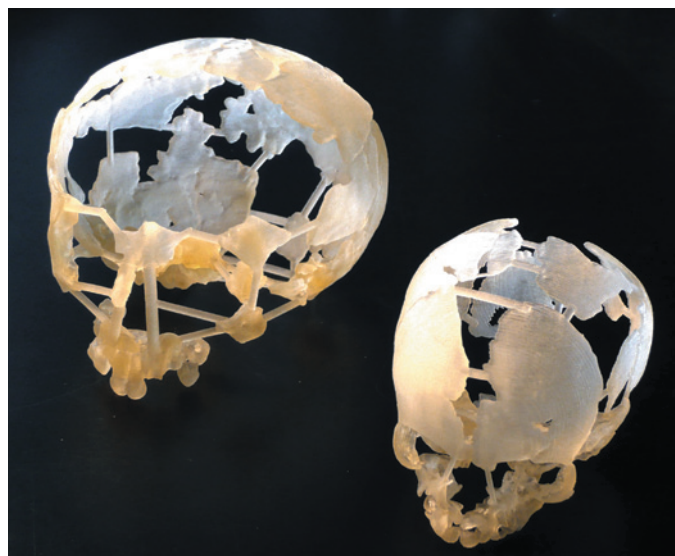
Olson is now trying to meld the tactile advantages of 3D printing with computer power: he has tagged printed models with small paper labels that can be recognized by a webcam, to create an ‘augmented reality’ view. In this way, a user can play with a physical model, while at the same time using the computer to explore aspects such as the potential energy of a given molecular arrangement. Olson is also looking forward to using printers that can more easily swap between rigid and bendable materials, so as to better replicate molecular behaviour such as protein folding.

THE CELLULAR MATRIX

Printer ‘inks’ aren’t limited to plastic. Biologists have been experimenting with printing human cells — either individually or in multi-cell blobs — that fuse together naturally. These techniques have successfully produced blood vessels and beating heart tissue. The ultimate dream of printing out working organs is still a long way off — if it proves possible at all. But in the short term, researchers see potential for printing out 3D cell structures far more life-like than the typical flat ones that grow in a Petri dish.

For example, Organovo, a company based in San Diego, California, has developed a printer to build 3D tissue structures that could be used to test pharmaceuticals. The most advanced model it has created so far is for fibrosis: an excess of hard fibrous tissue and scarring that arises from interactions between an organ’s internal cells and its outer layer. The company’s next step will be to test drugs on this system. “It might be the case that 3D printing isn’t the only way to do this, but it’s a good way,” says Keith Murphy, a chemical engineer and chief executive of Organovo.

Other groups are using 3D printing of plastic or collagen to construct scaffolds on which cells can grow. Carl Simon, a biologist with the biomaterials group at the US National Institute of Standards and Technology in Gaithersburg, Maryland, says that the intricacies of scaffold shape can help to determine how cells grow, or how stem cells differentiate into different cell types. With 3D printing, researchers have a very controlled way to play with different scaffold configurations to see which work best. One problem, however, is that most 3D printers can produce details on the scale of only tens to hundreds of micrometres, whereas cells sense



Printouts of Neanderthal skulls from a child (left) and a neonate.

differences at the single-micrometre level. Top-quality printers can currently achieve 100-nanometre resolutions by using very short laser bursts to cure plastics, says Neil Hopkinson, an engineer who works with 3D printing at the University of Sheffield, UK, but this is “still very much in the lab”.

CUSTOM TOOLS

In the meantime, basic plastic 3D printers are starting to allow researchers to knock out customized tools. Leroy Cronin, a chemist at the University of Glasgow, UK, grabbed headlines this year with his invention of ‘reactionware’ — printed plastic vessels for small-scale chemistry (M. D. Symes *et al. Nature Chem.* 4, 349–354; 2012). Cronin replaced the ‘inks’ in a \$2,000 commercially available printer with silicone-based shower sealant, a catalyst and reactants, so that entire reaction set-ups could be printed out. The point, he says, is to make customizable chemistry widely accessible. His paper showed how reactionware might be harnessed to produce new chemicals or to make tiny amounts of specific pharmaceuticals on demand. For now, other chemists see the idea as a clever gimmick, and are waiting to see what applications will follow.

Researchers in other fields have found a more immediate use for the technology. Philippe Baveye, an environmental engineer at Rensselaer Polytechnic Institute in Troy, New York, uses 3D printing to make custom parts for a permeameter — a device used to measure the flow of water through soils. Although commercially available devices are fine for routine work, he has often had to design his

own for more precise research — a task that previously required many hours on a lathe. Printing, he says, is much easier.

Perhaps more importantly, Baveye can share his product just by publishing the design file. “The idea of being able to reproduce experiments described in the literature is taking on a new meaning,” he says.

Others agree that the real power of 3D printing lies in its ability to put science into the hands of the many. Cronin wants to enable anyone — whether in the far corners of Africa or in outer space — to print their own tiny drug factory. Museums can already distribute exact copies of rare or delicate fossils as widely as they wish. And students can print out whatever molecule they’re trying to come to grips with. “Through 3D printing,” says Olson, “the ability to make physical models has become democratized.” ■ [SEE EDITORIAL P.6](#)

Nicola Jones is a freelance reporter based near Vancouver, Canada.

THE REAL POWER OF
3D PRINTING LIES IN ITS
ABILITY TO PUT SCIENCE INTO
THE HANDS OF THE MANY.

NATURE.COM

To see a 3D printer in action, go to:
go.nature.com/egidxx

UNLOCKING THE BRAIN

Much of our neural circuitry is fixed during childhood. Researchers are finding ways to unglue it, raising hopes for treating many brain disorders.

BY JON BARDIN

Growing up in the suburbs of New York City, Takao Hensch learned German from his father, Japanese from his mother and English from the community around him. “I was always wondering,” he says, “what is it that makes it so easy to learn languages when you’re young, and so hard once you begin to get older?”

Today, as a neuroscientist at Boston Children’s Hospital in Massachusetts, Hensch is at the forefront of efforts to answer that question in full molecular detail. Language acquisition is just one of many processes that go through a ‘sensitive’ or ‘critical’ period — an interval during development when the neural circuits responsible for that process can be sculpted, and radically changed, by experience (see ‘Open and shut’). During critical periods, children can make rapid progress at discerning facial features that look like their own, recognizing spoken language and locating objects in space. But within a few months or years, each window of opportunity slams shut, and learning anything new in that realm becomes difficult, if not impossible.

Or maybe not. What Hensch and others in the small, but rapidly advancing, field of critical-period research are finding is that those windows can be prised back open. “For the first time, we are beginning to understand the biology that underlies critical periods,” says Hensch. And that understanding is suggesting ways to intervene in various neural disorders, including intractable conditions such as adult amblyopia, in which information from one eye is not correctly processed by the brain, and possibly even autism. The work could even lead to ‘plasticity pills’ that enhance learning or help to wipe out traumatic memories.

“What’s so interesting about Takao’s work is that he has shown that even if you miss these critical periods, you still may be able to go back in and fix things,” says Charles Nelson, a

neuroscientist at Boston Children’s Hospital, who studies the developmental effects of early social deprivation on orphans in Romania. “The idea that you could intervene later and make up for lost time is compelling.”

The first scientist to popularize the notion of a developmental critical period was the Austrian biologist Konrad Lorenz, whose pioneering work in animal behaviour earned him a share of the 1973 Nobel prize. In the 1930s, Lorenz showed that if he took on the role of a mother goose within a few hours after goslings hatched, the baby geese would follow him as though he were their mother until adulthood. He called this process ‘imprinting’.

DOGMA, INHIBITED

The first scientists to explore the neural basis of a critical period were David Hubel and Torsten Wiesel, neurophysiologists at Harvard Medical School in Boston who carried out work on the visual system in the early 1960s. First they discovered that in the adult brain, many cells in the visual cortex respond to signals from only one eye. Then they showed that in kittens that had had one eye sutured shut, individual cells that normally would have fired in response to the closed eye instead responded to the open eye, eventually causing amblyopia¹. Shutting the eye of an adult cat did nothing, indicating that cells in the visual cortex were programmed during a key developmental window in the first few months of life.

Hubel and Wiesel lacked the tools to analyse how this programming worked at the molecular level, but they earned a Nobel prize in 1981 for their discovery. Their findings also inspired Hensch, during the 1980s, to change his undergraduate major from computer science and artificial intelligence to neurobiology. “Hubel and Wiesel’s work made me realize that there was

just so much we didn’t know about the actual biology of the brain,” he says.

Hensch got a chance to learn more when he began his PhD work in Michael Stryker’s neuroscience laboratory at the University of California, San Francisco (UCSF). Stryker’s group, like most researchers in the field, studied the critical period of the visual system as a model of critical periods in general, and had published a series of papers hinting at a new approach to understanding it.

For years, researchers had assumed that the brain’s ‘plasticity’, or its ability to learn during critical periods, was the work of excitatory neurons, which encourage neighbouring neurons to fire. But Stryker’s work suggested some kind of involvement by inhibitory interneurons, brain cells that dampen activity in their neighbours. Stryker’s team had found that, in kittens, a drug that increases inhibition during the critical period made the visual cortex resistant to Hubel and Wiesel’s trick: many neurons in that region began to fire in response to the closed eye rather than the open one².

Hensch followed up on this work in a collaboration with Michela Fagiolini and her colleagues at the RIKEN Brain Science Institute in Wako, Japan. The researchers looked at the critical period in genetically engineered mice that had slightly reduced levels of γ -aminobutyric acid (GABA), an inhibitory neurotransmitter.

The effect of that reduction was far greater than either Hensch or Stryker had imagined: whereas control mice went through a typical critical period and developed amblyopia when one eye was blocked, mice with GABA deficiencies did not develop amblyopia, or have a critical period at all. Hensch and his colleagues were able to restore plasticity by administering a benzodiazepine, a drug that enhances the inhibitory effect of GABA (ref. 3).

Inhibition, the authors concluded, was a

► NATURE.COM

Read more about critical periods in higher cognition: go.nature.com/hh9qil



hidden force driving the onset of the visual critical period. “At the time, these ideas were just so counter-intuitive,” Hensch says. “We were turning dogma on its ear.”

CLEVER MECHANISMS

Researchers have since begun to clarify the workings of this force. In 2008, Hensch and Alain Prochiantz, a neuroscientist at the Collège de France in Paris, found that when mice first open their eyes after birth, a protein called OTX2 is transported through the optic nerve from the retina to the visual cortex — a marathon in cellular distance.

In the visual cortex, the accumulation of OTX2 sparks a series of events that causes PV interneurons, inhibitory cells that contain parvalbumin (PV), to mature and trigger the beginning of the visual critical period. But this transport takes place only after visual input is

received; in mice raised in the dark, no OTX2 arrives in the cortex and no critical period ensues⁴. “I think this is a pretty clever mechanism from nature,” says Hensch, “because you don’t want to be plastic until you know that the periphery is functional and signals are coming in.”

But it was unclear how the PV interneurons triggered the critical period. An important clue came from a group led by Stryker with Arturo Alvarez-Buylla and Sunil Gandhi, also at UCSF. The researchers transplanted embryonic cells that were destined to become interneurons into the brains of young mice, says Alvarez-Buylla, after which the mice “started having two critical periods”. There was the typical critical period, caused by the mouse’s own interneurons, and then a later one, triggered when the transplanted interneurons began to mature⁵.

The transplanted cells, says Stryker, were

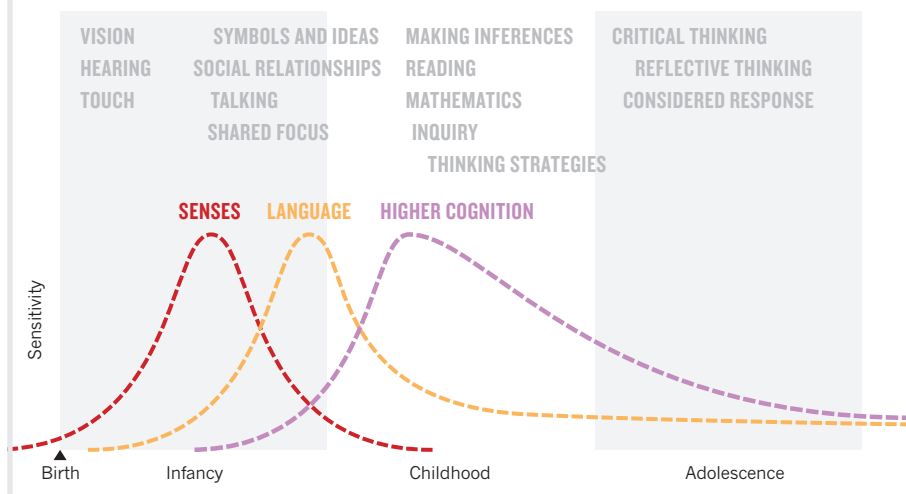
pushing the system’s ‘reset’ button. In the cerebral cortex of the adult brain, information travels through neural circuits along well-defined paths carved out by mature interneurons, which strongly inhibit some cells and not others. But in the transplant experiment, the maturing interneurons were making numerous weak connections with the older cells and inhibiting all the cells equally, overriding the brain’s previously defined circuits.

Only as those new cells matured were their connections pruned and strengthened, eventually carving out new permanent circuits. The findings suggested that the same mechanism — PV synapse proliferation followed by pruning — underlies all critical periods.

Hensch and others have found that critical periods do not just taper off as PV interneurons mature. Instead, they are shut down as the

OPEN AND SHUT

The human brain's sensitivity to learning seems to crest in three broad waves. The critical periods for cortical regions devoted to vision and other senses (red) open in infancy, then close tightly. Those for language (yellow) and higher cognition (purple) open later, and never close entirely. The successive waves allow a child to acquire increasingly complex skills (grey text).



brain slams on 'plasticity brakes' — presumably as a way to protect the newly optimized brain circuits from disruption by further input.

Hensch separates these 'brakes' into two categories — structural and functional. The first comprises physical structures such as the perineuronal net (PNN), a complex of macromolecules that attach to PV interneurons around the time that a critical period ends, and that seem to restrict the extent to which a neural circuit can change. Chemical breakdown of the PNNs in adult rats makes their brains prone to being rewired⁶.

Functional brakes are chemical compounds such as Lynx1 — a molecule, identified by Hensch and his colleagues, that shifts the balance of excitation and inhibition in the cortex by dampening the effect of the neurotransmitter acetylcholine. Experiments in mice show that the amount of Lynx1 in the brain increases at the end of the critical period, and its removal from adult brains, like degradation of the PNNs, seems to restore neural plasticity⁷.

Hensch says that what he finds particularly compelling about functional brakes is that they are relatively easy to release. One example of this is a behavioural intervention developed by Roger Li and Dennis Levi, optometrists at the University of California, Berkeley, for adults with amblyopia.

People develop amblyopia when problems such as cataracts or crossed eyes disrupt input to one of their eyes during early childhood, often leaving them without three-dimensional (3D) vision. The condition is considered untreatable once the critical period has ended. But when Li and Levi got people with amblyopia to play 40–80 hours of video games with their good eyes patched, most of them reported substantial improvements in visual function⁸. Describing one subject who was born with crossed eyes and had never seen the world in

depth, Li says: "Once she found out she was able to see some 3D, she immediately began to cry."

Hensch suggests that playing video games releases some of the brain's functional brakes. He notes that heightened attention, which often goes along with video-game playing, has been shown to increase the activity of acetylcholine — a surge that would counteract the damping effect of Lynx1.

WINDOWS OF OPPORTUNITY

Researchers have begun experimenting with drugs to reopen the critical period. Hensch and David Hunter, an ophthalmologist at Boston Children's Hospital, received approval in May to begin a phase I clinical trial for treating amblyopia with a drug that increases the amount of acetylcholine in the brain.

A similar study⁹, published in 2010 and led by neuroscientist Michael Silver of the University of California, Berkeley, found that when people with normal vision are given a drug that increases acetylcholine levels, they show greater improvements in visual acuity than people given a placebo. And a group led by Lamberto Maffei, a neurobiologist at the Scuola Normale Superiore in Pisa, Italy, has begun phase II clinical trials for amblyopia with selective serotonin re-uptake inhibitors, a class of drugs often used to treat depression.

Such research makes it easy to imagine pills or shots that could aid recovery from a severe brain injury, for example, or make it easier to learn a new language or forget a terrifying memory. Lifting plasticity brakes might even be useful in treating complex disorders such as autism, says Hensch. He points to the difficulty children with autism have integrating input from multiple senses at once — when looking at a person's facial expressions while

listening to them speak, for example. Such integration may require the critical periods for each sense to have occurred in a specific developmental sequence. "I think that autism is a good example of what can go wrong when these different sensory critical periods are mistimed," he says — a view for which there is some experimental evidence¹⁰.

For now, however, when it comes to the neural basis of complex psychiatric conditions such as autism, the experimental evidence is limited. But if tests could be created to identify risk factors for some developmental disorders, says Hensch, physicians may one day be able to deploy biologically informed interventions during the critical period, taking advantage of the brain's plasticity to set development on the right course.

But no one in the field is suggesting that the brain's critical periods should be tampered with casually. "When you reopen a critical period, there is, of course, always the possibility of a worse outcome," says Alvaro Pascual-Leone, a neurologist at Harvard Medical School, pointing out that disorders such as amblyopia occur because of harmful input during the original critical period.

And structural brakes are considerably more difficult to release than functional ones. In 2009, for example, researchers showed that chemically destroying PNNs in mice makes it easier to erase their fear memories, suggesting a potential treatment for conditions such as post-traumatic stress disorder¹¹. But to do this in humans could cause widespread brain damage that would outweigh any benefits. After all, says Hensch, the mechanisms that the brain uses to shut down critical periods are very complex, and they require a substantial amount of energy, "which gives us a good sense that they've evolved for a reason".

Stryker sounds a further note of caution. "I think it's a romantic notion that you can replicate the critical period later in life," he says. "Some things just don't unhappen." ■

Jon Bardin is a freelance writer based in New York.

- Wiesel, T. N. & Hubel, D. H. *J. Neurophysiol.* **26**, 1003–1017 (1963).
- Reiter, H. O. & Stryker, M. P. *Proc. Natl Acad. Sci. USA* **85**, 3623–3627 (1988).
- Hensch, T. K. *et al. Science* **282**, 1504–1508 (1998).
- Sugiyama, S. *et al. Cell* **134**, 508–520 (2008).
- Southwell, D. G., Froemke, R. C., Alvarez-Buylla, A., Stryker, M. P. & Gandhi, S. P. *Science* **327**, 1145–1148 (2010).
- Pizzorusso, T. *et al. Science* **298**, 1248–1251 (2002).
- Morishita, H., Miwa, J. M., Heintz, N. & Hensch, T. K. *Science* **330**, 1238–1240 (2010).
- Li, R. W., Ngo, C., Nguyen, J. & Levi, D. M. *PLoS Biol.* **9**, e1001135 (2011).
- Rokem, A. & Silver, M. A. *Curr. Biol.* **20**, 1723–1738 (2010).
- Rubenstein, J. L. R. & Merzenich, M. M. *Genes Brain Behav.* **2**, 255–267 (2003).
- Gogolla, N., Caroni, P., Lüthi, A. & Herry, C. *Science* **325**, 1258–1261 (2009).

SOURCES: T. HENSCH; MCCAIN, M.N.; MUSTARD, J.F. & MCCUAG, K. EARLY YEARS STUDY 3 CH. 2 (MARGARET & WALLACE MCCAIN FAMILY FOUNDATION, 2011).

COMMENT

POLITICS A call for a world without the danger of nuclear weapons **p.30**

BOOKS *Nature* regulars share their tips for holiday reading **p.34**

PROTEST Canadian scientists who blockaded coal train call for more direct action **p.38**

OBITUARY Phillip Vallentine Tobias, giant of hominin work, remembered **p.40**

COREIS



President John F. Kennedy meets the US Executive Committee of the National Security Council in October 1962 to discuss Soviet nuclear missiles in Cuba.

Decisions at the brink

Linguistic analysis reveals how advisers influenced President Kennedy during the Cuban missile crisis 50 years ago, argues **David R. Gibson**.

This October is the 50th anniversary of the Cuban missile crisis, when the United States and Soviet Union stared each other down after Soviet nuclear missiles were discovered on the island of Cuba.

Much has emerged about these events as successive troves of information have been uncovered — including records from the Soviet secret service and ruling Presidium¹, and from the US Department of State and Navy². Perhaps the most remarkable are the hours of audio recordings that US President John F. Kennedy made secretly of the deliberations of the Executive Committee of the National Security Council (ExComm), with which he met almost daily during the height of the crisis, between 16 and 28 October 1962.

The existence of the ExComm recordings — more than 20 hours' worth from these 13 days alone — was known only to Kennedy's close associates until they came to light during the Watergate investigations in 1973. The tapes were eventually declassified and, with a few excisions, released to the public in the mid-1990s. Historians have looked to them and the transcripts³ for insight into the positions taken by the dozen or so ExComm members, including Robert Kennedy, the attorney general and president's brother, and Robert McNamara, the Secretary of Defense at the time⁴.

I am the first to subject the recordings to scientific scrutiny, using the tools of conversation analysis. This subfield of sociology starts

with the meticulous transcription of conversation — with all its interruptions, slips and hesitations. Assuming that everything that happens does so for a reason, it asks why, when and with what consequences people do things such as interrupt, restart sentences and repeat others' words. It also helps to shed light on why particular discussions unfold as they do, from conversations about mundane matters, such as about what to have for dinner, to talk of greater moment.

The Cuban missile crisis is often held up as a model of rational decision-making. In many accounts, the only decision maker that mattered was Kennedy, and the ExComm's contribution is minimized. My study of these recordings, however, suggests that ►

CUBAN MISSILE CRISIS TIMELINE

How events unfolded during the height of the crisis in October 1962.



14 OCTOBER

Soviet missiles photographed in Cuba.

16 OCTOBER

President Kennedy informed; ExComm secretly convened.

22 OCTOBER

Kennedy announces discovery of missiles and imposition of naval blockade.

23 OCTOBER

Kremlin orders ships carrying weapons to be diverted from Cuba. Organization of American States supports blockade, lending it important legitimacy.

24 OCTOBER

US Strategic Air Command moves to DEFCON 2 (alert status just short of imminent nuclear war). Premier Khrushchev rejects blockade.

25 OCTOBER

Bucharest and *Völkerfreundschaft* allowed past blockade; photographs dramatically unveiled to United Nations Security Council.

26 OCTOBER

Marucla (Lebanese freighter under Soviet charter) intercepted. Khrushchev makes first proposal: to remove missiles in return for US pledge to not invade Cuba.

27 OCTOBER

Khrushchev makes second proposal, demanding withdrawal of Jupiter missiles from Turkey. U-2 plane shot down over Cuba. Kennedy publicly accepts Khrushchev's first proposal and secretly promises to remove Jupiter missiles.

28 OCTOBER

Khrushchev accepts.



► Kennedy was susceptible to persuasion, which was shaped by conversational vicissitudes and exigencies, yet also motivated and constrained by the pace of events in the outside world. My analysis challenges familiar ideas: that Kennedy steadfastly resisted calls for military action; that once he expressed a preference his advisers promptly fell in line; that his decisions were the obvious ones given the objective logic of the geopolitical situation; and that the ExComm's consideration of the options was thorough and balanced. In fact, the recordings reveal a president clearly swayed by the advice he received — and the manner in which he received it.

TIME TO TALK

The president learned of the missiles on 16 October (see 'Cuban missile crisis timeline'). His first major decision was to impose a naval blockade on Cuba, which he announced to the world on 22 October after revealing the discovery of the missiles. According to most accounts, President Kennedy was drawn to the idea of a blockade because it seemed like a strong action that did not involve an immediate military attack. Certainly the president was not eager to begin dropping bombs, but the blockade option itself carried serious risks. No one thought it would be sufficient to force Soviet Premier Nikita Khrushchev to withdraw the missiles already on the island, which meant that an air strike might eventually be needed to destroy them. In the interim, the Soviets would have time to finish work on the missiles, some of which could be fired at US cities — perhaps by accident or without authorization — in the event that the United States attacked.

How, then, did Kennedy come to choose a blockade? I identified and examined every instance of talk about a blockade and subsequent air strike in the tapes to see whether the danger of a missile launch was reiterated, omitted or anticipated but unsuccessfully articulated. I found a marked shift, early on 18 October, from frequent warnings, particularly from McNamara, about the danger of bombing operational missiles on the island, to a string of statements (the first from Llewellyn Thompson, former US ambassador to the Soviet Union) supporting the blockade and air strike but omitting any reference to this danger.

After this shift, whenever someone tries to reintroduce the risk of a nuclear response, that person is interrupted, talked over or ignored. Twice, for instance, Robert Kennedy tries to warn about the danger of allowing the Soviets time to finish work on the missiles before a US attack, but both times he is thwarted. The first time, Maxwell Taylor, Chairman of the Joint Chiefs of Staff, talks over him. The second time, McNamara repeatedly interrupts him before he can finish his thought.

I believe such efforts at 'suppression'

reflected a growing determination to reach consensus on some course of action when the best course remained elusive, and to combat attempts to threaten this consensus. This in turn allowed President Kennedy to make a choice, which he did on the 20 October, that he could justify in light of recent talk — even if it flew in the face of much of what had been said earlier.

This consensus was possible because, for the first week of the crisis, the United States kept its knowledge of the missiles a closely guarded secret, so the ExComm could revisit the blockade option again and again until objections were squelched. When all options look bad, a decision may hinge on the ability of groups to talk in circles until such a time as the more frightening consequences of one course of action disappear from view.

OVERTAKEN BY EVENTS

Kennedy's second main decision concerned the implementation of the blockade. Soon after the president made his public announcement on 22 October, Khrushchev ordered Soviet ships carrying weapons to be diverted away from Cuba² (although a warhead-laden ship called the *Alexsandrovska* slipped into port in Cuba before the blockade was put into effect). Other ships were allowed to continue on their course. Although Kennedy believed that these ships were not carrying missiles, some felt that an interception would be needed so that Khrushchev would not doubt US resolve.

Four Soviet-bloc ships were put forward as candidates. Two (the *Kimovsk* and *Grozny*) were ordered by Khrushchev to turn back before they could be intercepted. Another (the East German *Völkerfreundschaft*) was a passenger ship that no one had the stomach to challenge for fear of loss of life. The only serious candidate for interception was the *Bucharest*, a Soviet tanker that, lacking deck cargo, was almost certainly carrying only oil. The ship turned up at the blockade line early on the morning of 25 October and, before it could be discussed by the ExComm, was permitted to pass after it declared its cargo benign. Thus, the ExComm meeting later that morning was, in part, about whether to chase it down and board it belatedly.

The standard story is that the president let the *Bucharest* go because he wanted to give Khrushchev more time to back down, or to respond to diplomatic overtures at the United Nations³.

In fact, the tapes reveal a Kennedy who is very worried about appearing weak if he does not order the ship to be boarded. His hand is stayed by opposition from his more-vocal advisers. They assure him that the failure to intercept could easily be justified to domestic and international audiences, even as they thwart attempts to give sustained consideration to the way that the Soviets

BETTMANN/CORBIS

HULTON-DEUTSCH COLLECTION/CORBIS



Aerial shots of Soviet missile launch sites in Cuba provoked a naval blockade by the United States.

might interpret the move. Kennedy remains unconvinced. By the end of the meeting he still seems inclined to board the ship, but postpones a decision until later in the day.

Before the ExComm reconvened, however, a state-department briefing for legislators leaked the fact that the *Bucharest* had already crossed the blockade. Rather than admit indecision, the Pentagon announced that the navy had let it pass after ascertaining that it carried no contraband. When the ExComm reconvened later in the afternoon, it took the passage of the tanker as irreversible — showing how indecision can be transformed into decision by the pace of events.

Kennedy's third big decision concerned the deal that resolved the crisis. Late on 26 October, he received a private letter in which Premier Khrushchev offered to withdraw his missiles in return for a US pledge not to invade the island. Before this could be discussed by the ExComm the next morning, the group received another offer, this one made publicly, demanding that President Kennedy remove the North Atlantic Treaty Organization's nuclear (Jupiter) missiles from Turkey. The ExComm was dumbfounded, but Kennedy's advisers quickly converged to recommend that the president accept the first offer and ignore the second. The president resisted, however, insisting that Khrushchev would never be content with such a deal, having publicly proposed a missile trade.

The group reached an impasse on the afternoon of 27 October. Neither side could muster an argument that the other side would accept, even though everyone felt pressured to respond to Khrushchev's offer before the end of the day, partly for fear that if there was any delay he would issue additional demands.

On the tapes, President Kennedy's advisers try, in various ways, to work around their

boss. Twice, for instance, the president is allowed to talk on at length about the futility of accepting Khrushchev's first offer, receiving no challenge despite providing openings (such as pauses) for disagreement.

Several times people even suggest that President Kennedy remove himself from the letter-drafting process. Robert Kennedy does so most forthrightly. This results in an explosion of laughter at his audacity, and at the way he is pointing out the elephant in the room — that the president has become the primary impediment to consensus as to what the president should do.

Finding little support for the position he has taken, President Kennedy eventually relents and agrees to the wording of a letter that accepts Khrushchev's first offer, even though it hints at later negotiations over "other armaments".

Yet Kennedy remained unconvinced that this would be enough. Following a more intimate meeting off-tape on that night, he deployed his brother to secretly deliver a promise that the Jupiter missiles would be removed from Turkey within a few months of the end of the crisis, on the condition that Khrushchev did not publicize that part of the deal.

Khrushchev accepted, having been about to agree to the terms of Kennedy's letter when he received word of the further concession². This marked the beginning of the end of the crisis. Although there followed several months of tense negotiations about inspections and Soviet nuclear bombers¹, by Christmas all the offending weapons had been removed from Cuba.

NOT SO RATIONAL

The Cuban missile crisis had a good ending, and the ExComm deliberations deserve some credit. But this was not a rational decision-making process in any traditional

sense. The discussions were undoubtedly beneficial in bringing to light the various options and their associated risks, and in forcing the different factions (the 'hawks' and 'doves') to hone their arguments. But they were also structured by conversational machinery — based, for instance, on the rules of turn-taking and the expectation that one say something relevant to whatever was said last — that is not easily put to the service of comparing the consequences of competing courses of action.

In fact, the need to reach consensus at each stage required the ExComm to avoid, or cease, consideration of some of the risks: the risk of having to bomb operational missiles if the blockade failed; the risk that letting the *Bucharest* past would leave Khrushchev with the impression that Kennedy was weak; and the risk that by accepting Khrushchev's first offer and ignoring his second, the first real path out of the crisis (like-for-like missile withdrawal) would be sacrificed.

The same process could have had a different ending. Had a consensus in favour of the blockade been forged even a day sooner, for instance, it might have gone into effect in time for the Navy to intercept the *Aleksandrovsk*. Given that this ship was under orders to sink itself rather than be boarded²,

"The same process could have had a different ending."

that would have sent history careering down a different path entirely.

The lesson for scientists, advisers and policy-makers is that the details and mechanics of conversation matter. Talk is useful for decision-making, but its conventions do not ensure that sustained attention is given to all the things that could go wrong. Given enough time, all branches of the decision tree might receive their due, but during a crisis time is in short supply, and the hardest decisions might require that some branches are neglected and even wilfully abandoned. ■

[SEE COMMENT P.30](#)

David R. Gibson is a lecturer of sociology at Princeton University in New Jersey. His book *Talk at the Brink: Deliberation and Decision During the Cuban Missile Crisis* (Princeton University Press) is published this month.

e-mail: drigibson@princeton.edu

1. Fursenko, A. & Naftali, T. *One Hell of a Gamble: Khrushchev, Castro, and Kennedy, 1958–1964* (Norton 1997).
2. Dobbs, M. *One Minute to Midnight* (Knopf, 2008).
3. May, E. R. & Zelikow, P. D. *The Kennedy Tapes* (Harvard Univ. Press, 1997).
4. Stern, S. M. *Averting the 'Final Failure': John F. Kennedy and the Secret Cuban Missile Crisis Meetings* (Stanford Univ. Press, 2003).
5. Kennedy, R. F. *Thirteen Days: A Memoir of the Cuban Missile Crisis* (Norton, 1969).



ILLUSTRATION BY PETE ELLIS/DRAWOOD.COM

A call for global nuclear disarmament

Danger from nuclear weapons is mounting. It is time to take control of the nuclear fuel cycle and move towards a world without warheads, says **Scott D. Sagan**.

Declassified documents have revealed that the 1962 Cuban missile crisis was far more dangerous than anyone knew at the time. An American U-2 spy plane accidentally flew into Soviet airspace and US fighter jets armed with nuclear-tipped missiles entered the Bering Strait to rescue it. US Minuteman missile controllers jury-rigged their systems so that they could launch the nuclear missiles on their own if necessary. Pentagon planners began preparing for the possible invasion of Cuba, totally unaware that Soviet tactical nuclear weapons were already deployed on the island and that local commanders had the authority to use them¹. US President John F. Kennedy and Soviet Premier Nikita Khrushchev avoided nuclear war in October 1962, but we now know how close they came to disaster² (see page 27).

Fifty years on, we live in a nuclear world that has not just two superpowers but nine nuclear-weapon states, with new ones looming on the horizon (see 'World of weapons'). The governments of these emergent nuclear states may not make the same mistakes that Russia and the United States made during the cold war, but they will make others.

We have entered a grave new world where governments believe that shielding themselves with their nuclear weapons will allow them to engage more safely in aggressive action, and increase nuclear proliferation by selling their technology to other governments. And it is a world where nuclear materials and weapons are becoming

increasingly vulnerable to theft and use by terrorists.

Looking at this prospect, some politicians and analysts have optimistically argued that new nuclear powers will behave cautiously and that a stable form of global nuclear deterrence is likely. For example, in 2007 Jacques Chirac, then president of France, told the press that it "would not be very dangerous" if Iran obtained nuclear weapons. If Iranian leaders ever used the bomb, he argued, Tehran would be destroyed immediately in retaliation³. Others are more pessimistic and insist that the only way to counter nuclear proliferators such as Iran is through preventive military operations. But such attacks are unlikely to be completely effective, could trigger wider wars and, over the long term, could actually

NATURE.COM
More on scientific
steps to nuclear
disarmament:
go.nature.com/sxcyne

encourage the spread of nuclear weapons.

Neither passivity nor preventive war are likely to lead to a safer nuclear future. Given the gravity of the risks we face, careful and steady movement towards global nuclear disarmament should be our goal. The complex and global nature of emerging nuclear dangers will require complex and global solutions. Scientists, engineers and governments must work together to improve verification of nuclear disarmament and to strengthen international control of uranium enrichment and plutonium reprocessing technologies.

EMERGING DANGERS

The dangers of the emerging nuclear landscape can be seen clearly in Pakistan. Within months of testing its first nuclear weapon in May 1998, the Pakistani military sent soldiers disguised as *mujahedin* guerillas into Indian-held Kashmir, a move that sparked the 1999 Kargil War and led to a series of nuclear threats between New Delhi and Islamabad. Pakistani nuclear scientist Abdul Qadeer Khan then established an infamous network that sold centrifuge technology to Iran, North Korea and Libya. Pakistan also highlights the risks of nuclear terrorism: the army is fighting internal Pakistani Taliban threats and has been penetrated by Islamist radical insiders⁴.

Islamabad is aware of these dangers and has strengthened security measures, some with US assistance, including steps to ensure the reliability of its personnel and to protect its nuclear arsenal from terrorists while the weapons are stored in bunkers at military bases⁵. Yet a dangerous 'vulnerability–invulnerability paradox' remains. In a future crisis with India, the Pakistani military is likely to place its nuclear arsenal on alert, mating the warheads with mobile missile launchers and moving the missiles outside its bases and into field positions. This operation would make Pakistan less vulnerable to attack by India, but make the weapons far more vulnerable to seizure by terrorists.

Similar nuclear dangers can be seen in Iraq, which tried and failed to develop the bomb. We now know, from documents captured after the fall of Baghdad in 2003, that Saddam Hussein abandoned his secret nuclear-weapons-development programme after UN inspectors discovered it in the wake of the first Gulf War. But we can learn important lessons about how new nuclear-weapons states might act by studying how Saddam envisioned using nuclear weapons as a shield for aggression towards other nations.

The captured documents reveal that before the invasion of Kuwait in 1990, Saddam predicted that Iraq would have nuclear weapons within five years. He asked his colleagues, rhetorically, "If the Arabs were to have a nuclear bomb, wouldn't

they take the territories that were occupied after 1967?"⁶ In the late 1970s, Saddam told Iraqi leaders that owning nuclear weapons would permit Iraq to launch a conventional war against Israel without the fear that Tel Aviv would, in desperation, retaliate with its nuclear arsenal. The transcript of Saddam's secret speech is chilling: "We can guarantee the long war that is destructive to our enemy, and take at our leisure each metre of land and drown the enemy with rivers of blood."⁷ The world is fortunate that Saddam was forced to destroy his nuclear programme after the 1991 Gulf War.

North Korea poses a serious nuclear threat. Its government has behaved aggressively since conducting its first nuclear test in 2006, and a second in 2009. In 2010, it was accused of sinking a South Korean Navy vessel, the *Cheonan*, which killed 46 seamen. North Korean artillery shelled the South Korean island of Yeonpyeong in November that year, killing two marines and two civilians. In addition, North Korea is such an economic basket case that it seems willing to sell almost anything: from counterfeit currency, fake pharmaceuticals and illicit drugs to the more lucrative business of smuggling missile and nuclear technology. North Korean leaders may not sell actual nuclear bombs to other countries — they have only enough nuclear material for a handful of weapons for their own country's use — but they have shipped missiles without nuclear warheads to Pakistan and Iran. They were also caught secretly selling uranium hexafluoride (used to make nuclear reactor fuel, or a bomb) to Libya in 2004 and a plutonium production reactor to Syria in 2007.

IRAN'S NUCLEAR SHIELD

Why should we worry about a nuclear Iran? The gravest danger is not that leaders in Tehran will order a suicidal nuclear first strike on Israel or the United States. The real danger is that if Iran acquires nuclear weapons, leaders in Tehran will see them as a shield behind which they can engage in conventional and terrorist aggression.

Iran is already a major arms supplier to Hezbollah, the Lebanon-based Shi'a militia, and supports its attacks against civilian and military targets inside Israel and other Middle Eastern countries. Iran also covertly armed Iraqi Shi'a militia fighting US troops in Iraq. Moreover, the Islamic Revolutionary Guard Corps — responsible for running Iran's operations in support of terrorist organizations — controls the nuclear programme. To have nuclear command and terrorist ties within the same organization is a deadly mix. Saudi officials have declared that if Iran gets the bomb, they too will develop a nuclear arsenal, which would add one more to the list of dangerous new nuclear states.

Given how far Iran has already gone with

its nuclear programme, some politicians and analysts advocate air strikes — or even a preventive war — against the country. An attack, however, would have serious global economic costs, cause many Iranian civilian casualties and would be unlikely to solve the problem. Preventive attacks can invigorate rather than destroy a covert nuclear-weapons programme. We learned this in 1981, when Israeli jet fighters bombed Iraq's only nuclear reactor, and Saddam then started a better-funded and more covert nuclear-weapons

programme based on secret uranium-enrichment facilities⁸.

Fortunately, we are not yet at the crisis point when the only options are to attack Iran or live with a nuclear-armed Iran.

The global policy objective should be to push the day of reckoning over the distant horizon. We must create time for sanctions and cyber-sabotage to delay Iran's progress, raise the costs of its unwillingness to constrain its enrichment programme, and enable diplomatic initiatives to be effective. A potential negotiated settlement could include agreement for regular international inspections but allow Iran to keep the Natanz uranium-enrichment centre and a store of low-enriched uranium. This would not be ideal because it would leave Iran with a long-term nuclear break-out option. But it would be better than a preventive attack or blind faith in nuclear deterrence.

Nuclear weapons may have been a dangerous necessity to ensure the cold war stayed cold. But scholars and policy-makers who are nostalgic for the brutal simplicity of that era's nuclear deterrence do not understand how much the world has changed. The choice we face today is not between a nuclear-weapons-free world or a return to bipolar cold war deterrence, but between a world free of nuclear weapons or one with many more nuclear states.

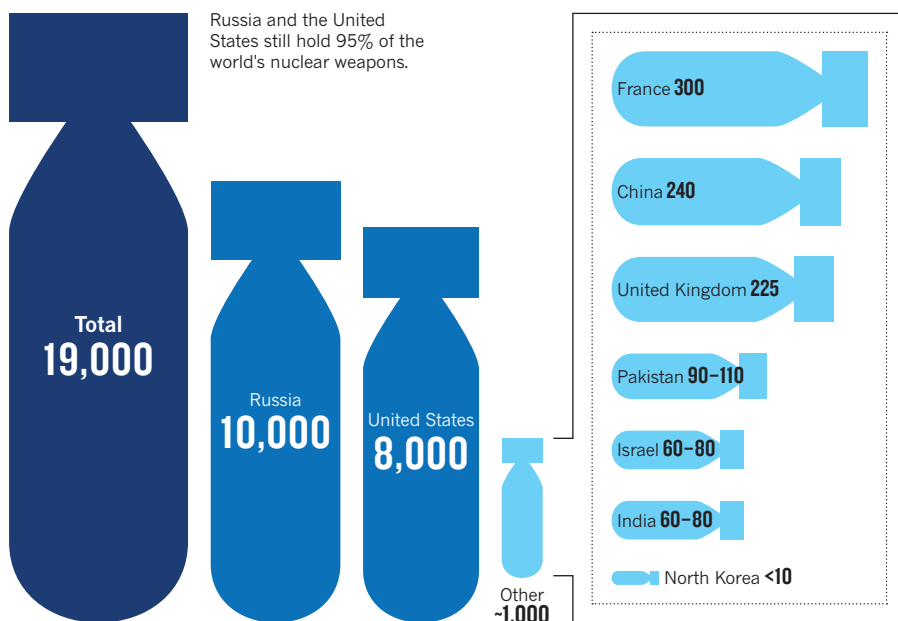
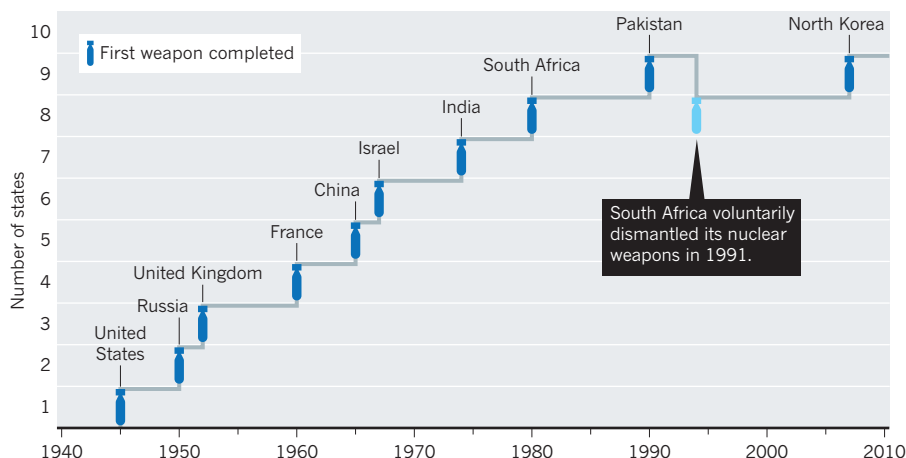
ROAD TO DISARMAMENT

The technical and political challenges that confront proponents of nuclear disarmament are complex and serious. We lack adequate disarmament-verification technology, such as techniques to permit remote sensing of covert weapons-related activities. Some allies rely on extended nuclear security guarantees — US commitments to retaliate with nuclear weapons if the ally is attacked. And increasing numbers of countries are both developing nuclear power and demanding the right to build uranium enrichment and plutonium reprocessing facilities, which could be used for peaceful or nefarious purposes. It is by no means clear that the United States and

"The choice is between a world free of nuclear weapons or one with many more nuclear states."

WORLD OF WEAPONS

Since the start of the cold war, the number of nuclear states has risen to nine, adding to regional tensions.



other nuclear-weapons states will overcome these challenges any time soon. What is clear is that existing nuclear-weapons states cannot disarm without the partnership of non-nuclear-weapons states⁹.

President Barack Obama, unlike previous US presidents, has correctly noted that the nation's membership in the Non-Proliferation Treaty (NPT) commits it to working towards nuclear disarmament¹⁰. By taking its NPT obligations seriously, his administration's 2010 Nuclear Posture Review argued, "we strengthen our ability to mobilize broad international support for the measures needed to reinforce the non-proliferation regime and secure nuclear materials worldwide"¹¹. At the successful 2010 NPT Review Conference, most non-aligned countries praised the New START (Strategic Arms Reduction Treaty) US–Russian arms-control agreement, and supported the creation of nuclear-power inspection protocols. This shows that

progress on disarmament aids success in non-proliferation¹².

It will be more difficult to achieve disarmament if there are many 'latent nuclear-weapons states' with their own uranium enrichment or plutonium reprocessing facilities that could produce fuel for nuclear power or weapons. It is imperative that governments work together to establish international control of such enrichment and reprocessing technologies — although they are used to produce nuclear fuel for power reactors, they could also be misused to build nuclear bombs. All such facilities should, in the future, be managed by an international agency and built under permanent safeguard agreements that ensure that if any country were to withdraw from the NPT, it could not then legally use the facility for a nuclear-weapons programme.

There are many important tasks for scientists and engineers on what will be, at best, a long and winding road towards nuclear

disarmament. We will need advanced verification technologies that can detect covert weapons-related activities remotely. As states move down the disarmament path, improved stockpile-stewardship programmes will be needed to ensure that nuclear arsenals remain reliable as they shrink, without the need for nuclear testing. We will also need to design advanced nuclear reactors that can produce energy for civilian purposes, but with less risk that this will lead to weapons proliferation or theft of nuclear materials by terrorists.

A nuclear-weapons-free world will not be a world without conflicts of interest or war. Nor will it be a utopia in which governments never feel tempted to cheat on their global obligations. Indeed, the maintenance of 'global zero' would require that conventionally armed major powers be prepared to enforce nuclear disarmament and non-proliferation commitments in a fair and vigorous manner. Potential proliferators may have to be 'forced' to comply.

The strategic challenges we face are daunting and we may end up with small nuclear arsenals rather than attain the global-zero landmark. But even that would be a much safer world than the one we live in now. If we fail to work together to achieve nuclear disarmament, the world we are heading towards — bristling with nuclear-weapons states, with more nuclear weapons, and the ever-present threat of nuclear terrorism — is even more fraught with danger. ■ [SEE COMMENT P.27](#)

Scott D. Sagan is the Caroline S. G. Munro professor of political science and a senior fellow at the Center for International Security and Cooperation, Stanford University, California 94305, USA. e-mail: ssagan@stanford.edu

1. Dobbs, M. *One Minute to Midnight* 191–192, 258, 276–279 (Knopf, 2008).
2. Sagan, S. D. *The Limits of Safety* 81–91 (Princeton Univ. Press, 1993).
3. Sciolino, E. & Bennhold, K. Chirac strays from assailing a nuclear Iran. *The New York Times* (1 February 2007).
4. Lister, T. & Kassim, A. Arrest of Pakistani officer revives fears of extremism within military. *CNN World* (22 June 2011).
5. Sanger, D. E. *The Inheritance: The World Obama Confronts and the Challenges to American Power* 220–224 (Harmony, 2009).
6. Brands, H. & Palkki, D. *Int. Security* **36**, 133–166 (2011).
7. Palkki, D., Stout, M. E. & Woods, K. M. *The Saddam Tapes* 223–224 (Cambridge Univ. Press, 2011).
8. Braut-Hegghammer, M. *Int. Security* **36**, 101–132 (2011).
9. Sagan, S. D. *Shared Responsibilities for Nuclear Disarmament: A Global Debate*. AAAS Occasional Paper, 1–13 (AAAS; 2010).
10. Miller, S. E. et al. *Nuclear Collisions: Discard, Reform & the Nuclear Nonproliferation Regime*. AAAS Occasional Paper, 1–41 (AAAS; 2012).
11. *Nuclear Posture Review Report* (US Dept. of Defense, 2011); available at www.defense.gov/npr.
12. Müller, H. *The Nonproliferation Review* **18**, 219–236 (2011).

SOURCES: S. D. SAGAN AND REV. POL. SCI. **14**, 225–244 (2011); *NUCLEAR STOCKPILE REPORT*; R. S. NORRIS & H. M. KRISTENSEN BULL. AT. SCI. **66**, 77–83 (2010).



SUMMER BOOKS

With the annual exodus from labs and lecture theatres on the horizon, *Nature's* regular reviewers and editors share some gripping holiday reads.

EDITORS' PICKS



Survival of the Beautiful: Art, Science and Evolution

DAVID ROTHENBERG
Bloomsbury: 2011.

What to do about beauty? Charles Darwin worried about the excess of the peacock's tail, but he invoked "endless forms most beautiful" to affirm the truth of his evolutionary theory.

Beauty is integral, not incidental, to the evolutionary process, says philosopher David Rothenberg in his rich account of the ways in which creatures make art. Evolution produces beauty as well as practicality. Moths, squid, elephants, proteins and birds produce individual as well as generic patterns.

Look at the variety of structures made by bowerbirds. Satin bowerbirds decorate their designs with blue items (plastic spoons will do). Golden bowerbirds use "two kinds of flowers: fresh olive-green ones and dried cream-coloured blossoms".

The resultant artwork is an end in itself, as well as a medium for reproductive behaviour, Rothenberg argues. Sexual selection requires flaunting, display, theatre, extravagance, scents and song. It demands excess, beyond what is necessary for survival.

Yet Rothenberg seeks laws that underlie performance in art and in science. He argues that human exposure to abstract art has helped us to find new patterns in scientific domains.

Gillian Beer is professor emeritus of English literature at the University of Cambridge, UK.



Everyday Information: The Evolution of Information Seeking in America

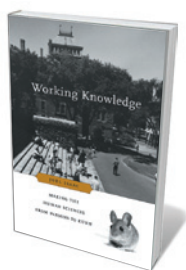
EDITED BY WILLIAM ASPRAY
AND BARBARA M. HAYES
MIT Press: 2011.

Everyday Information is an important book for anyone who has wondered how we got through life before the Internet. Without a web browser or smart phone, how did we find information? Why did we trust it? And what have global information networks altered?

For transactions such as purchasing a car or buying aeroplane tickets, the authors explain, we have switched from relying on local, trusted networks (such as travel agents) to depending on distant information networks (such as magazines and websites). But not all local networks are trusted: no one likes used-car salesmen.

We increasingly consume and create online information, sharing our interests in fantasy sports or comic books. Vibrant subcultures can result, and the Internet is changing who takes part. Online ecologies have restored female comic-book readership, for example, outside the male enclaves of specialist shops. So the demise of the corner shop has a silver lining. The book's US focus invites comparative studies.

Thomas Misa is director of the Charles Babbage Institute at the University of Minnesota in Minneapolis.



Working Knowledge: Making the Human Sciences from Parsons to Kuhn

JOEL ISAAC
Harvard University Press: 2012.

Unlike physics, chemistry and biology, which took on their modern forms in the nineteenth century, the social sciences coalesced only during the twentieth. The tale of their consolidation, rise and subsequent slide is often narrated as a clash of ideologies: scientific versus humanistic. In *Working Knowledge*, historian Joel Isaac reveals how institutional circumstances shaped the field.

He does so by putting its pioneers, including sociologist Robert K. Merton, psychologist B. F. Skinner and philosopher of science Thomas Kuhn back into the contexts in which they learned their crafts. He explores Harvard University in Cambridge, Massachusetts, where each spent formative periods.

Isaac documents brilliantly how they made their ways on the margins of departments. Elders of the university aimed to restrict specialization, so rising fields such as psychology and sociology were pursued in informal, interdisciplinary groups.

Isaac's elegant study shows how debates over method spring from efforts to embed new types of inquiry in the classroom.

David Kaiser is professor of the history of science at the Massachusetts Institute of Technology in Cambridge.



1493: How Europe's Discovery of the Americas Revolutionized Trade, Ecology and Life on Earth

CHARLES C. MANN.
Granta: 2011.

European exploration and exploitation of the new world in the fifteenth century changed the world. Journalist Charles Mann relates how the 'Columbian Exchange' altered cultures and transformed much of the Americas, Asia and Africa into ecological copies of Europe.

The book's scope is vast. It ranges from US tobacco to Bolivian silver, the collapse of the Chinese and Spanish economies as a result of commodity-exchange feedback loops, and malaria's role in the creation of the United States.

When humans move, they take food and drugs with them. Plants form the matrix in which the rest of life operates, so their exchange has profound and often unanticipated consequences. The introduction of the potato to Europe, for example, led to over-dependence on it, resulting in famine, mass emigration and political and economic consequences that seem far from the humble spud.

Mann demonstrates the paradoxical nature of the ecological effects linking the world. Globalization is nothing new.

Sandra Knapp is a botanist at the Natural History Museum in London.



Biomedical Computing: Digitizing Life in the United States

JOSEPH NOVEMBER
The Johns Hopkins University Press: 2012.

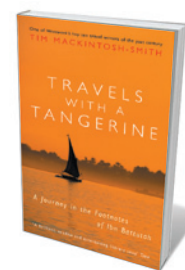
Computers changed research in the life sciences in the 1950s and 1960s. Historian Joseph November engagingly relates how. The shift was far from inevitable, but was partly a result of a deliberate effort by the US National Institutes of Health (NIH).

The effort was inspired by Robert Ledley, a dentist turned operations researcher turned computer specialist whom James Shannon, then director of the NIH, brought in to overcome hostility to computing.

The life sciences also changed computers. November argues that the Laboratory Instrument Computer (LINC) was the first personal computer. This was developed for biomedical research by Wesley Clark and his colleagues at the Massachusetts Institute of Technology (MIT) in Cambridge. In 1963, the NIH funded researchers to take LINC's back to their labs, provided that they spent six weeks that summer at MIT. Participants helped to assemble their LINC's, which were behind schedule in production.

November is now working on a biography of Ledley. That is good news: November's style is convincing and compelling.

Paula Stephan is professor of economics at Georgia State University in Atlanta, and author of *How Economics Shapes Science* (2012).



Travels with a Tangerine: A Journey in the Footnotes of Ibn Battutah

TIM MACKINTOSH-SMITH
Picador: 2003, reissued 2012.

The Tangerine in the title is Ibn Battutah, who was born in Tangier, Morocco, in 1304. Between 1325 and 1354, he travelled three times as far as Marco Polo, following the Silk Road across central Asia and visiting India, Indonesia, the Philippines, China and Zanzibar. He also crossed the Sahara Desert to Timbuktu, an oasis city emerging as a centre of scholarship that would rival Paris and Oxford.

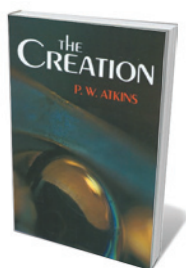
Arabist Tim Mackintosh-Smith was so enthralled by those journeys that, 650 years later, he retraced Ibn Battutah's first great trip to Mecca. This book is a delightful account of his experiences. Two further volumes span Ibn Battutah's later journeys.

Ibn Battutah was insatiably curious and wrote extensively on ethnography, geography and botany. In 1348, he provided one of the best contemporary accounts of the Black Death.

I first 'met' him 50 years ago on the Malabar Coast of India, which he described, including mention of black sand (containing ilmenite and monazite ores). This was helpful to my research on the effects of exposure to high local background radiation.

I highly recommend Mackintosh-Smith's trilogy.

Robin Weiss is professor of viral oncology at University College London.



The Creation

PETER ATKINS
Freeman: 1981.

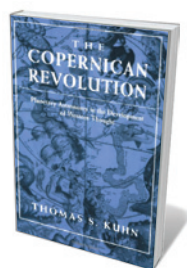
In 1981, chemist Peter Atkins summed up evolution by natural selection: “Once molecules have learnt to compete, and to create other molecules in their own image, elephants, and things resembling elephants, will in due course be found roaming through the countryside.”

That is one of the more verbose moments in Atkins’s elegant, laconic and consequently short book, *The Creation*. In it, he relates how a creator who produced the world’s infinite variety of life forms could have been “infinitely lazy” — so the act of creation itself needs no explanation. The accumulation of changes arising from unguided movements within the space of possible forms can produce objects whose complexity baffles the best scientists.

But it gets even better. To update Atkins, once evolution creates a human mind in which ideas compete, tools such as spear throwers, computers and space shuttles will also, in due course, appear. Humans conduct tournaments of natural selection among ideas, so cumulative cultural adaptation will produce objects of ever-greater complexity. No light bulbs need switch on in our minds.

We could have been as ‘lazy’ as Atkins’s creator, and still iPads and things resembling iPads would have been destined to be part of our future.

Mark Pagel is professor of evolutionary biology at the University of Reading, UK.



The Copernican Revolution

THOMAS S. KUHN
Harvard University Press: 1957.

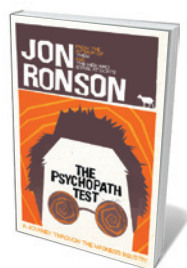
Before philosopher Thomas Kuhn debuted the idea of paradigm shifts in *The Structure of Scientific Revolutions*, he wrote *The Copernican Revolution*, published in 1957, the year when Sputnik went into orbit. Reading this book in the current age of extrasolar planets, genetics and string theory is eye-opening.

More than 500 years ago, Nicolaus Copernicus displaced Earth from the centre of the Universe, beginning profound changes in scientific thought. After following the thread of astronomical minutiae that led to that bold move, Kuhn argues that Copernicus didn’t foment a revolution himself, but rather hit on a way to mesh the world view of the past with the astronomy of the future. The true revolution swelled slowly, bursting onto the world centuries later.

Kuhn’s tale resonates with modern discoveries about planetary systems other than our own. The book opens our minds to how differently the history of astronomy could have played out had we lived on a world with twin suns, on a moon around a Jupiter-like planet or in a system packed with gas giants looping inside our planet’s orbit.

As we learn about other worlds, Kuhn’s analysis reminds us that we are poised for our own scientific revolution.

Caleb Scharf is director of astrobiology at Columbia University in New York.



The Psychopath Test

JON RONSON
Picador: 2011.

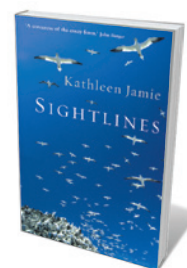
Several years ago, I interviewed a psychologist who studied people who believed that they had been abducted by aliens. She argued that, although deluded, they were not crazy.

I found myself wondering about the science behind labels of madness, as journalist Jon Ronson does in *The Psychopath Test*. Ronson delves into the madness industry to explore this question, travelling from a UK criminal psychiatric facility to New York, where he meets a former death-squad leader who says he just wants to be liked.

The Psychopath Test’s main subject is a checklist for psychopathy that is widely used by the criminal-justice system. It was developed by Canadian criminal psychologist Robert Hare. It has been attacked for leading to imprisonment of people who are deemed psychopaths, a label that is not included in the current edition of the *Diagnostic and Statistical Manual of Mental Disorders*. Ronson finds that the checklist criteria, such as ‘grandiose sense of self-worth’, can be applied flexibly.

Ronson’s gift is his ability to illuminate impenetrable worlds. Hare’s checklist is already the subject of criticism, but Ronson details many studies showing how bad psychiatry has been at diagnosing mental illness.

Sharon Weinberger writes on national security issues from Washington DC. Her most recent book is *A Nuclear Family Vacation* (2008).



Sightlines

KATHLEEN JAMIE
Sort Of Books: 2012.

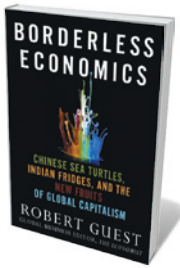
All too often, whatever our discipline, we are prone to tunnel vision. It takes a poet to stand back and show us the world as it is. *Sightlines*, Kathleen Jamie’s astounding series of essays, makes us look anew at the familiar and the strange.

From the whirling madness of a Scottish gannetry, Jamie narrows her focus to a pod of orca whales, circling silently below the cliffs where she is perched. In a hospital pathology lab, she surprises the clinical consultant with her desire to see other species — not dolphins, but “the bacteria that can pull the rug from under us” as she peers through an electron microscope.

Jamie visits Bergen, Norway, to see the Whale Hall museum under restoration. She joins the team scrubbing the leviathans’ bones with toothbrushes. Sitting in the ribcage of a blue whale is “like being in a very strange taxi, caught in traffic”.

Hers is a vital sense of being. She takes us from a teenage stint as an archaeologist digging up the long dead, to her triumph as she makes it to the abandoned Scottish archipelago of St Kilda and tells of its vanished inhabitants, who built “small dark closets, just to get some seclusion, some corrective to the sky, the sea and wind and each other”. Rejoicing in Jamie’s economy of words, we are never less than alive, bursting out of our cells and into the light.

Philip Hoare is a writer based in Southampton, UK. His latest book is *Leviathan* (2008).



Borderless Economics

ROBERT GUEST
Palgrave Macmillan: 2011.

Humans have been spreading around the globe for economic reasons for 50,000 years, since the earliest people in Africa chased animals for food. Now, movement across borders is more restricted than ever, yet this basic economic pursuit continues because disparity among nations is at a historical peak.

When individuals migrate to pursue opportunity, the places they leave behind and those they reach all gain. Mobile people bring cheap labour, consumers and dynamism. They send back money and information; some return physically.

In *Borderless Economics*, journalist Robert Guest has written — with data, anecdotes, and humour — an optimistic account of the state of this age-old pursuit, adapted to the political, economic and technological possibilities of today.

He travels from the United States to closed North Korea to explore immigrants' economic adaptation. Guest sheds light on concepts such as micro-multinationals: small-scale collaborations that span the globe, perhaps starting with a few university friends who source technologies in one country, serve markets in another and manufacture products in a third.

This is splendid book, but could have given more attention to diasporic low-skilled workers.

Iqbal Quadir is founder and director of the Legatum Center at the Massachusetts Institute of Technology in Cambridge.



Victorian Science and Literature Part 1: Negotiating Boundaries

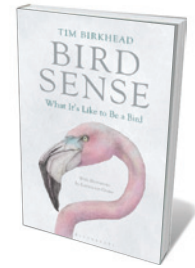
EDITED BY PIERS J. HALE AND JONATHAN SMITH
Pickering & Chatto: 2011.

How does scientific knowledge affect poetry? This question preoccupied many Victorian writers. *Negotiating Boundaries*, the first in an eight-volume edition of writings on Victorian science and literature, shows scientists, poets, novelists and critics in collaboration and passionate competition.

Poetry, wrote priest John Henry Newman in 1858, “is always the antagonist to science”. As science advances, poetry recedes. Nonsense, riposted geologist Hugh Miller: only feeble poets are put off by knowledge. For a “great poet”, science is not an enemy, but a “devoted friend”.

The Victorians used poetry to grapple with science. Scientific and literary writers tried to broker a relationship between fact and emotion. Photochemist and statistician Robert Hunt, for example, wrote in *The Poetry of Science* (1848) that “to rest content with the bare enunciation of a truth, is to perform but one half of a task”. The other half is to communicate that truth, surrounding it with impulses of feeling that “pass from soul to soul”. Science needs poetry, just as much as poetry needs science.

Alice Jenkins is professor of English literature at the University of Glasgow, UK.



Bird Sense: What It's Like to be a Bird

TIM BIRKHEAD
Bloomsbury: 2012

Tim Birkhead is more than your average twitcher. His fascinating book attempts to put us inside birds' eyes, ears, minds and even hearts, in seven chapters covering sight, hearing, touch, taste, smell, magnetic sense and emotions.

A behavioural ecologist, he specializes in sperm studies and avian infidelity; his chapter on touch focuses almost indecently on bird sex. He takes us on adventures to find his subjects in the wild and rehabilitates little-known but prescient nineteenth-century naturalists.

Birkhead offers irresistible snippets. In addition to the conventional visible spectrum, many birds see the world in ultraviolet; wading birds can taste the presence of worms in wet sand; robins can “see” Earth's magnetic field through their right eyes but not their left.

Birkhead generally avoids anthropomorphizing birds, although he comes close to crossing this line in the chapter on emotions. I know the feeling, having watched penguins in Antarctica hug their mates.

And although we will never really know how it feels to be a bird, Birkhead still leaves us with astonishing insight into how they sense their own ways through the world.

Gabrielle Walker is a writer based in London. Her latest book is *Antarctica* (2012).

REVIEWS: GO.NATURE.COM/SOFVRK

Correspondence

Take direct action on climate inaction

We are scientists recently arrested in Canada for blockading a 125-car train carrying coal destined to release 26,000 tonnes of carbon dioxide into the atmosphere. We joined 11 other Canadians in this act, despite the personal risks and potential negative impact on our careers.

Time is running short and our dialogues on climate change with Canada's conservative government have been futile, which is why we undertook this extreme action. We were following the example of NASA climatologist James Hansen, who has been arrested three times in the past three years for civil disobedience in protesting against the mining of coal or development of the Canadian oil sands.

If the rate of carbon emissions does not decrease soon, the 2°C threshold for serious consequences of climate change could be broken this century (M. New *et al. Phil. Trans. R. Soc. A* **369**, 6–19; 2011). Yet many nations, including Canada and the United States, remain more concerned with building infrastructure to extract and transport fossil fuels than with seeking alternative energy solutions.

Civil disobedience has a long-standing tradition of inducing social change when those in power fail to act. Governments are neglecting their responsibility to future generations. Because science is built on professionalism and objective evidence, media coverage of our arrests will ensure that they, and the voting public, receive a forceful message.

Alejandro Frid *West Vancouver, British Columbia, Canada.*
alejfrid@gmail.com

Lynne Quarmby *Simon Fraser University, Burnaby, British Columbia, Canada.*



Doubt in Australia's emissions scheme

Policy uncertainty is dogging emissions-trading schemes. On 1 July, Australia introduced the largest carbon-pricing scheme outside the European Union, with a higher price for carbon permits than in the EU carbon market (F. Jotzo *Nature Clim. Change* **2**, 475–476; 2012).

When we surveyed a sample of Australian large emitters, carbon financiers and carbon-market experts (see go.nature.com/jlehiy), we found that 79% think there will still be a carbon price in 2020. But 38% expect that the current scheme will be repealed by the end of 2015, in line with the opposition's pledge. Of those who expect repeal, half think that a carbon price will be reinstated in Australia by 2020.

According to our survey, the average expected carbon price for the first three years of Australia's scheme is predicted to start near the 'fixed' (legislated) price of Aus\$23 per tonne of carbon dioxide equivalent. It then falls to an expected Aus\$11 per tonne for 2016 before rising to Aus\$22 per tonne in 2025.

Assessments of future prices vary greatly between experts, indicating pervasive uncertainty. Low-carbon investments depend on expectations about future prices, and can be hampered by

too much uncertainty. Managing prices in emissions-trading schemes could help. One way would be to use a fixed price, as in Australia, or a price band. Another would be to vary the supply of permits, as proposed for the EU carbon market, in which prices have dropped following economic troubles.

Setting carbon prices in line with domestic policy ambitions may be an attractive option for other countries, including for China's planned emissions-trading schemes. It will not alleviate policy uncertainty, but it can reassure businesses that low-carbon investments will pay off financially and make revenues more predictable.

Frank Jotzo *Australian National University, Canberra, Australia.*
frank.jotzo@anu.edu.au

Spanish cuts: careers come to abrupt end

Carmen Vela suggests that the 22.5% cut to the already limited Spanish science budget is an opportunity for improvement (*Nature* **486**, 7; 2012). But her optimism is unrealistic.

Spain already has substantially fewer researchers per capita than other members of the European Union such as France or Germany. Entire grant programmes have disappeared

and important research institutes have laid off many highly qualified scientists. Even the Ramón y Cajal programme for young researchers has failed to fulfil its contractual obligations, forcing an end to the careers of bright young scientists.

This all comes at a time when research and development offers Spain the only certain route of recovery from economic collapse.

Sergio P. Acebrón *German Cancer Research Center, Heidelberg, Germany.*
s.perezacebron@dkfz-heidelberg.de

Spanish cuts: more economic damage

The Spanish government's secretary for research, development and innovation has attempted to put a positive spin on the country's severe shortfall in research funding (C. Vela *Nature* **486**, 7; 2012). But the latest drop in funds is unlikely to make the little science that remains more competitive.

Returns from investment in science are unpredictable. But limited funding will mean that high-risk projects get left out and that scientists will be driven abroad or choose alternative careers. Investment during the bonanza years will have served no purpose.

These deep cuts for science will deprive the Spanish economy as a whole. Creating an environment that is conducive to research and innovation calls for a long and sustained input from government, irrespective of economic cycles.

Manuel Corpas *The Genome Analysis Centre, Norwich, UK.*
manuel.corpas@tgac.ac.uk

Spanish cuts: reform bureaucratic culture

The Spanish government's draconian cuts to its science budget do indeed present us with an opportunity (C. Vela

Nature **486**, 7; 2012): to overhaul and reform our entire research system. We are then more likely to be in a competitive position when the economic situation improves.

Unnecessary bureaucracy should be a prime target. The cumbersome administrative regulations of public institutions too often conflict with researchers' needs and hold up their work.

For example, the Spanish Research Council (CSIC), which accounts for 16% of Spain's publications (according to Thomson Reuters' ISI Web of Science), could short-circuit administrative delays by allowing its CSIC Foundation to manage its research projects. This increased flexibility would also save public money.

Budgets are set to increase by 58% in Europe's Horizon 2020 research-funding programme. Spanish researchers should ensure that they are in line for the extra funding, despite the national cuts.

Miguel B. Araújo *National Museum of Natural Sciences, CSIC, Madrid, Spain.*
maraujo@mn.cn.csic.es

More funding for studies of ageing

To enable ageing populations to stay healthy, gerontology research needs to be allocated more resources and given greater priority. This will improve the quality of life for elderly people and relieve the economic burden of their long-term sickness.

Studies of ageing processes could help to prevent and treat chronic-disease states that are common after the age of 65, such as cardiovascular disease, type 2 diabetes or neurodegenerative syndromes. For instance, the rate of increase in cardiovascular disease in people aged 50–90 is 7-fold in men and 16-fold in women. Arguably, this makes ageing a bigger risk factor than elevated cholesterol (*Incidence and Prevalence: Chart Book on Cardiovascular and Lung Diseases* NHLBI, 2006). Yet funding for gerontology research

is paltry compared with that for, say, cancer or HIV.

The US National Institutes of Health (NIH) Geroscience Interest Group, which involves experts from 20 NIH institutions, has taken a step in the right direction in setting up funding mechanisms for cross-cutting areas of research.

Gerontology research also needs its own funding advocates. The American Association for Retired Persons in Washington DC and similar non-profit groups in other countries should take the lead. Scientists, too, can change the prevailing negative attitude towards ageing in their communications with government, colleagues and the public.

Brian Kennedy *Buck Institute for Research on Aging, Novato, California, USA.*
bkennedy@buckinstitute.org

Trade threat could be even more dire

Manfred Lenzen and colleagues suggest that 30% of species globally may be threatened by the international trade in commodities (*Nature* **486**, 109–112; 2012). We feel that this estimate is too conservative.

The scale and magnitude of damage to biodiversity resulting from the recent sharp rise in trade volume will probably take decades to realize (F. Essl *et al. Proc. Natl Acad. Sci. USA* **108**, 203–207; 2011). Exports worldwide increased by 119% between 1990 and 2011 (see go.nature.com/au5chf), notably from 'megadiverse' countries such as China, Brazil, India and Indonesia. These countries are also major exporters of commodities such as biofuel, which has a huge detrimental impact on tropical and subtropical biodiversity.

The real threat to biodiversity is underestimated by analysing only globally threatened species, because these are largely outnumbered by regionally threatened species (M. Winter *et al. Proc. Natl Acad. Sci. USA* **106**, 21721–21725; 2009).

Both of these considerations

need to be included in attempts to quantify biodiversity damage from international trade. For example, although the Forest Stewardship Council certification process does include the impacts of trade, it does not consider time lags that may worsen biodiversity damage (www.fsc.org).

Franz Essl *Environment Agency Austria, Vienna, Austria.*
franz.essl@umweltbundesamt.at
Marten Winter *Helmholtz Centre for Environmental Research — UFZ, Halle, Germany.*
Petr Pyšek *Institute of Botany, Academy of Sciences of the Czech Republic, Průhonice; and Charles University, Prague, Czech Republic.*

Support for a cholera vaccine stockpile

As researchers and policy-makers in Rwanda's health sector, we congratulate Guinea and Haiti on integrating oral cholera vaccination into their regular control efforts during epidemics (see *Nature* <http://doi.org/h2c>; 2012). Rwanda has also learned valuable lessons about the costs of inaction from the world's delayed response to cholera outbreaks in post-earthquake Haiti.

Because cholera is endemic in the neighbouring Democratic Republic of Congo, Rwanda's health workers are trained in outbreak detection and management. Immediate action meant that a small cholera outbreak in the Nkamira refugee camp, currently at four times normal capacity, was promptly contained.

Nine patients at the camp presented with severe watery diarrhoea within a few days of each other (29 May to 3 June 2012). These cases were rapidly confirmed as cholera by laboratory testing; all were successfully treated with intravenous solution, oral rehydration salts and ciprofloxacin antibiotic. Patient contacts were traced and monitored, and sanitation campaigns reinforced. No further cases have been reported.

Within 24 hours, preparations were made to procure Shanchol

cholera vaccine for all 10,000 camp residents. However, it was not necessary to deploy this vaccine because traditional methods had worked in time.

We support the call for a global stockpile of cholera vaccine and argue for rapid integration of prevention and care at the first signs of a cholera outbreak.

Agnes Binagwaho *Ministry of Health, Kigali, Rwanda.*
agnes_binagwaho@hms.harvard.edu
Thierry Nyatanyi *Rwanda Biomedical Center, Kigali, Rwanda.*
Cameron T. Nutt *Dartmouth Center for Health Care Delivery Science, Hanover, New Hampshire, USA.*
Claire M. Wagner *Harvard Medical School, Boston, Massachusetts, USA.*

Extend ethnicity of human microbiome

The Human Microbiome Project has generated 5,177 taxonomic profiles of microbes sampled from 242 people, mostly westerners (*Nature* **486**, 207–214; and *Nature* **486**, 215–221; 2012). It would be enlightening to extend these studies to a wider range of human populations to see how their microbial profiles vary. For example, the diversity and function of organ flora in humans could be affected by such factors as geography, food habits, environment, age, traditions and changing lifestyles.

Bhaswati Chatterjee *Centre for Cellular and Molecular Biology (CCMB) Quarter, Nacharam, Hyderabad, India.*

Suman S. Thakur *Centre for Cellular and Molecular Biology, Hyderabad, India.*
sst@ccmb.res.in

CORRECTION

The Outlook article 'The lost appetites' (*Nature* **486**, S16–S17; 2012) contained two inaccuracies. The panda has not lost its sweet receptor, only the umami receptor. And goitricin can worsen hypothyroidism, not prevent hyperthyroidism. The sentence that follows has been changed to reflect that.

Phillip Vallentine Tobias

(1925–2012)

Palaeoanthropologist who pioneered description of African hominins.

Phillip Vallentine Tobias, known to his friends and colleagues as PVT, was the doyen of the palaeoanthropological community. His descriptions of the early hominin fossils found at Olduvai Gorge, Tanzania, are part of the bedrock on which rests our knowledge of human origins.

For half a century, he stewarded the excavations at Sterkfontein, a group of fossil-rich caves northwest of Johannesburg, South Africa. He was a colleague of Louis and Mary Leakey, the husband-and-wife team who did much to establish that humans originated in Africa. And he remained in South Africa throughout his career — with all the professional complications and opportunities for activism that entailed. All this gave him a distinctive role at the University of the Witwatersrand ('Wits') in Johannesburg, and in the field.

Tobias was born and received his early education largely in Durban, South Africa. In the first volume of his autobiography, *Into the Past* (Picador Africa, 2005), he suggested that the premature death of his sister Val from diabetes, as well as his visits to the Durban Natural Science Museum, were the reasons for his interest in science.

A six-decade-long connection with Wits began when he started his BSc in physiology, histology and embryology, which he gained in 1946. While an undergraduate, Tobias met his mentor, Raymond Dart, who was famous for discovering *Australopithecus africanus*, the first early hominin to be unearthed in Africa. Tobias completed a medical degree in 1950, but opted for a career in research.

After attaining a PhD in genetics in 1953, Tobias took part in an expedition to study the San bushmen in the Kalahari Desert. This drew him into physical anthropology. He spent 1955 at the University of Cambridge, UK, examining hominin fossils curated in England and France. In 1956 he toured the United States, returning to South Africa late that year. Through these travelling fellowships he gained contacts that he nurtured for the rest of his career.

In 1959, shortly after Dart retired, Tobias was appointed as head of the anatomy department at Wits, a position he held until he retired in 1990. That was also the year that he got his big scientific break. Louis and Mary Leakey invited him to analyse the *Zinjanthropus boisei* cranium that Mary had just discovered at Olduvai Gorge (since

reclassified as *Australopithecus boisei*, then as *Paranthropus boisei*).

Tobias's collaboration with Louis and the British anatomist John Napier to describe the Olduvai fossils became the basis for the description of *Homo habilis* — a new, and at

early variety of *Homo*, and Clarke argued that others belonged to *Paranthropus robustus*. The Sterkfontein fossils are crucial to debates about whether hominins moved freely between eastern and southern Africa, or evolved independently in the two regions.

In 1994, Clarke discovered four australopithecine foot bones in a box of fossils that had been collected in 1977. He and Tobias published an account of the bones, and Clarke has since recovered much of the skeleton, nicknamed 'Little Foot'. The collection is on track to being the most complete early hominin skeleton ever recovered.

Tobias's undergraduate contemporaries at Wits' medical school included Sydney Brenner and Aaron Klug. Whereas their research interests led them away from South Africa, Tobias stayed put. He had to remain where the fossils were. He became a consistent thorn in the side of the South African apartheid regime. His vociferous opposition, in speeches and in demonstrations that were ruthlessly suppressed by the authorities, began in 1948 when he became president of the National Union of South African Students. He was no less ardent as a senior academic. Although his scientific prominence gave him a measure of protection, his moral and physical courage was not to be underestimated. It was distressing that Tobias was occasionally tarred with the apartheid brush and denied access to conferences in other countries.

Wedded to his work, Tobias never married. Apart from watching his beloved cricket at the Bidvest Wanderers Stadium in Johannesburg, he either worked or travelled. I visited him once at his home. Even though he owned three large tables, we had to eat dinner from trays on our laps, because his 'habilis volume' was spread out on all the other surfaces.

As meticulous about his manners and dress as he was about his writing and lectures, PVT was kind and encouraging to students and young researchers, me included. South Africa and the palaeoanthropological community are immeasurably poorer for his passing. ■

Bernard Wood is the University Professor of Human Origins at George Washington University, Washington DC 20052, USA. Tobias became his unofficial mentor after Wood visited Wits to study fossil craniums in 1972. His research builds on Tobias's seminal work on the Olduvai hominin fossil record. e-mail: bernardawood@gmail.com



first contentious, hominin species. A monograph, *The Cranium and Maxillary Dentition of Australopithecus* (Zinjanthropus) boisei (Cambridge University Press, 1967), followed. Two mammoth volumes entitled *The Skulls, Endocasts and Teeth of Homo habilis* (Cambridge University Press, 1991) provided exquisite details of subsequent hominin discoveries from Olduvai.

For much of his career, Tobias focused on excavations at the Sterkfontein caves. In 1958 Sterkfontein became the property of Wits, and Tobias ramped up the operations there, which continue to this day. Initially, the excavations were supervised by Alun Hughes, who had worked with Dart. Since 1991, they have been led by Ronald Clarke, who had assisted the Leakeys.

By the early 1990s, Tobias's team had collected more than 500 hominin fossils, mostly of *A. africanus*. However, some (including the jaw in his left hand, pictured) were judged by Tobias and Hughes to belong to an

AVUS/GALLO IMAGES/GETTY

CAREERS

TURNING POINT Cancer researcher wins photography contest **p.131**

CAREERS TOOLKIT Resources and expert advice on careers issues go.nature.com/mfgmok

NATUREJOBS For the latest career listings and advice www.naturejobs.com



CORBIS

advice to other scientists whose first language isn't English: "You just have to write the manuscript yourself. Lots of principal investigators will do the job for you, but that's not really good training." His persistence paid off: *Developmental Cell* published his manuscript.

When it comes to publishing in English-language journals, the philosophical differences between languages can be even trickier than the grammatical ones. Written English tends to be more direct and straightforward than Chinese, notes Cathy Wang, a retired forest-products professor based in Vancouver, Canada. She co-designed and teaches on the Scientific Writing in English for Chinese Authors distance-learning course at the National Pingtung University of Science and Technology in Taiwan. "In English, if you want to write well," she tells students, "you need to be clear, concise and forthright."

All English-as-a-second-language (ESL) scientists confront cultural differences, language barriers and grammatical peculiarities when trying to publish their work in international peer-reviewed journals written in English, which often hold the keys to successful scientific careers. Roughly written manuscripts might not even get sent out for review, so researchers must master not only writing in a foreign tongue, but also writing about technical, complex topics in a standardized way. (Several ESL scientists contacted for this story declined to comment, concerned that admitting to difficulties could have negative consequences for their careers.) And peer-reviewed publishing has its own 'culture' — rules about authorship, conflicts of interest and plagiarism that can differ from a scientist's own cultural norms.

THE RIGHT WORDS IN THE RIGHT ORDER

When everything — from following journals' instructions for submission to answering peer-review criticisms — is difficult, the path to publication can be frustrating and time-consuming. English's dominance as the language of science puts non-native speakers at a disadvantage from the start, says Judith Hertog, an ESL specialist at Dartmouth College's Institute for Writing and Rhetoric in Hanover, New Hampshire. Journal editors may not explicitly reject papers because of poor English, notes Dugald McGlashan, publisher of Asian Academic Journals for Nature Publishing Group in Melbourne, Australia, citing his own experience. But, he says, poor language does create a barrier that scientists must overcome to ►

PUBLISHING

Foreign tongues

Non-native English speakers face challenges when trying to publish. But there are resources that can provide help.

BY KENDALL POWELL

When drafting his first scientific paper in English, one Chinese third-year graduate student took the hard road. He insisted that his adviser, whose first language was English, allow him to write the paper himself. Although the student (who asked not to be named) had used

English textbooks at university in Beijing, and could read and understand English well, it was a monumental challenge to find the right words and craft sentences to express his scientific thinking.

"I went back and forth with my principal investigator maybe 20 times and it took four, five, maybe six weeks," recalls the researcher, now a biologist at a Colorado university. His

► get their papers accepted. “That’s a fine difference, but publishers cannot publish a paper that doesn’t make sense.”

There are strong pressures to publish in English. In some countries, universities offer cash rewards or other perks for getting papers into high-impact English-language journals, or even make it a requirement for career advancement (see *Nature* 463, 142–143; 2010).

Scientists who struggle do have options, from institutional support at university writing centres to services offered by private consultants and editing companies. Many research universities and government agencies now set aside funding specifically for language editing, to ensure that poorly written English does not impede publishing aspirations.

PROPER PREPARATION

Attention to language should start early, before the researcher begins to draft the paper, and continue until it is ready to be submitted (see ‘Correct, clear, concise’). First and foremost, researchers should read as many English-language papers in their field as they can, and pay careful attention to those that are easiest to read and understand.

“My most basic advice is to look for models of well-written papers,” says Hertog. She suggests that researchers keep a notebook for useful words and phrases, to help them to “master the language of their discipline”. Emulating an effective author is a great idea, she says, as long as no direct copying occurs.

Early-career scientists writing their first English-language papers could try enrolling



Biomedical editor Rashmi Nemade says scientists should learn how to write compellingly in English.

in a writing course, during which they might be able to construct and revise the paper from start to finish under an instructor’s guidance. Researchers should avoid relying too heavily on grammar-checking tools in word-processing programs, which aren’t designed for the nuances of scientific writing. And they should avoid translation software or online tools: the results can be barely readable, or even complete nonsense. If a scientist needs translation services, they should find a professional translator.

Another option is to use institutional writing resources. Many universities in the United States have writing centres that offer services specifically designed for international graduate students and postdocs. For example, Hertog gives free, private writing consultations and

holds weekly study meetings for ESL students.

She encourages ESL graduate students to begin writing about their research in English as early as possible. Advisers, she says, should assign research summaries, reviews of papers, poster presentations and other writing exercises in the first year of the graduate programme; even if they don’t, students should attempt such exercises for themselves.

Once a researcher has prepared the best version of a manuscript that they can, they should seek informal reviews from native English speakers. Morgan Tucker, a life-sciences and medical editor based in Boulder, Colorado, says that this step is imperative — even a few rough spots in the language and readability of a paper can prevent a journal editor from sending it out to reviewers. “There are too many unwritten rules in English. English speakers don’t really know the rules, but we can hear them,” he says. “There’s a cadence to the language that you have to hit to get the rhythm right.”

HIRED HELP

If there is only one English speaker in a lab or department, they can quickly be overwhelmed with requests to edit others’ work. Other editing options have sprung up in response to growing demand.

Some editing consultants specialize in particular fields of study. Such professionals can offer both English and scientific editing; that is, they can help to make the written arguments in a paper or a grant proposal as strong as possible. These services generally cost between US\$150 and \$450 per manuscript, depending on how much time the job takes. (Some authors choose to thank editors in the paper’s acknowledgements section, although there is no standard of practice.)

In addition to language editing, Tucker rewrites and reorganizes parts of papers, and highlights sections that do not make sense. He also formats the article according to the target journal’s guidelines, and compares it with a recently published paper. “I’ll make it look like it belongs in that journal to help clients get published,” he says.

Dan Csontos, founder of Elevate Scientific in Lund, Sweden, offers seminars, two-day workshops and editing services for ESL physicists and materials scientists. He also goes beyond language help, coaching authors on how to present their research logically and in the context of their field. He and his colleagues “can read any physics paper, understand the logical flow of ideas and see the gaps”, says Csontos, a former editor at *Nature Physics* and *Nature*.

Not all editors are specialized. Some publishers — including Nature Publishing Group — offer English-language editing services that cover a range of disciplines, and editing companies often offer a menu of translation, writing, editing and other publishing services.

WRITING TIPS

Correct, clear, concise

Similar glitches come up again and again for non-native English speakers writing scientific manuscripts in English. Here are a few ways to mitigate problems and frustrations.

● **Problem:** Basic language mistakes, including leaving out articles such as ‘a’ and ‘the’; confusion between words that sound the same when spoken, such as ‘site’ and ‘sight’ or ‘led’ and ‘lead’; and use of inappropriately informal language, such as ‘We interviewed 20 guys for our experiment’, or ‘This is a pretty accurate result’, in a paper.

Solution: Get a native English speaker to review the writing and provide feedback.

● **Problem:** The argument lacks a coherent or logical thread to take the reader from one point to the next.

Solution: Get a colleague from another field to read the paper and point out where it is confusing.

● **Problem:** Sentences are long and confusing, with subjects and verbs far apart.

Solution: Break up sentences whenever

possible: in general, they should be no more than 20 words long. Read through the text and reword to move subjects and verbs closer together where necessary.

● **Problem:** Transitions between sentences or paragraphs are abrupt or don’t flow well.

Solution: Establish the topic at the beginning of a sentence and provide new information at the end. Linking back to the previous sentence gives the reader ‘stepping stones’ of understanding. For example: ‘In India, the tiger population is making a comeback. This recovery is due to anti-poaching laws introduced in the 1990s. These laws also make it illegal to kill tigers for sport.’

● **Problem:** Using language that is too varied or elaborate, or that has a literary, rather than scientific, style.

Solution: Stick to one word or phrase for each concept — otherwise, the reader could get confused. The simplest language is usually the most readable. **K.P.**

Editage in Treves, Pennsylvania, has websites in English, Japanese, German, Portuguese, Korean, simplified Chinese (for mainland China) and traditional Chinese (for Taiwan). The Japanese version has a free 'chat' function for scientists to ask general questions about publication and authorship practices; this function may soon be expanded to other languages. For a 4,000-word manuscript, Editage's basic language editing starts at around \$200; its premium service costs about \$400 and takes the paper from start to finish. Editors ensure that the article complies with the target journal's author guidelines and provide help with cover letters and responses to reviewer comments. They also provide coaching on issues such as journal selection, authorship, disclosure of conflicts of interest and peer-review function in Western journals.

Edanz, an editing firm headquartered in Hong Kong, caters mainly for scientists from Asia, the Middle East and South America. It offers comprehensive editing; a web-based journal-selection tool that can use a paper's abstract to draw up a list of the top ten most appropriate target journals from a pool of more than 10,000; a mock peer-review service; and in-person workshops on how to structure manuscripts.

"We teach authors how to be persuasive and construct stories that meet reader expectations," says Benjamin Shaw, China and global director of Edanz in Beijing. The company charges a flat fee based on the length and difficulty of the paper; articles below 4,000 words usually cost \$200–250. For that price, they do the work necessary — including multiple revisions — to bring it up to the level for publication in an international journal, says Tom da Costa, managing director for the firm's Japanese operations, based in Fukuoka.

Shaw notes that high-quality editing companies should be willing to answer questions about pricing, about who is doing the editing and about security and privacy. Edanz publishes its editors' profiles online so that scientists know who will be previewing their research.

Regardless of whether they seek help from colleagues, university services, private consultants or large firms, ESL scientists have to consider how to 'sell' their research to editors and reviewers, says Rashmi Nemade, founder of BioMedText, a biomedical writing and editing firm in Columbus, Ohio. "It's got to be entertaining and compelling," she says. "This concept is sometimes lost on scientists in general, but particularly those from other cultures where work is all about merit." ■

Kendall Powell is a freelance writer based in Lafayette, Colorado.

TURNING POINT

Martin Oeggerli

Best known as a scientific photographer, Martin Oeggerli takes close-up images of microscopic creatures and structures that have been featured in scientific publications and art galleries worldwide. Although he is still a postdoc in cancer research at the University of Basel in Switzerland, his photography has become his passion and his main career track. Oeggerli received his third Best Scientific Image prize in March, in The EMBO Journal's annual cover contest, for capturing the most interesting microstructure.

How did you come to take up photography?

I drew a lot as a child — I especially liked snakes and lizards — and I always tried to draw as accurately and in as much detail as I could. When I started to study, however, I didn't have much time to draw, so in 2001, when I was 26, my father gave me my first camera, a digital Nikon Coolpix 995. In one month, I took 20,000 images of insects, snakes and lizards.

How did technology affect your photography?

That first camera had just 3.2 megapixels, but it had an outstanding macro function and excellent post-processing software. I looked at a close-up picture I had taken of a fly and thought it was spectacular. So I kept trying to get closer, and started to buy micro-adapters for the camera. Finally, I turned to scanning electron microscopy (SEM). These images are the best: there is huge depth of field, which is very appealing because it resembles the depth we see with our own vision. The tiniest things are magnified 500,000 times or more.

When did you first get public recognition?

In 2006, I entered an international photography contest in Germany and submitted an image of a pollen grain germinating, just at the point of releasing a pollen tube. The pollen grain hydrates as soon as it attaches to the female part of the flower, and can grow the tube in a couple of seconds. I had tried to distinguish the different structures and present them in a way that looked natural. I didn't expect an award, but I won third place in Best Research Image.

Has there been broad interest in your work?

Yes. I grew up in a small town and didn't go to many art galleries, but I was interested in books and magazines. I started to read *National Geographic* when I was 12 or 13. I always liked the pictures — it is regarded as one of the world's most important photographic magazines. So I couldn't believe it when *National Geographic* published my images in December 2009 as part



A. TRAINA

of a feature about pollen grains. I started scientific microscopy as a hobby and never expected it to go anywhere, but this was my international breakthrough. It confirmed to me that the work I do really is of interest.

What was your EMBO winning entry?

An image of a network of mosquito eggs. By trapping air, this mesh-like web repels water, allows the eggs to float and keeps the whole structure from sinking.

You tint your images. Why?

SEM is always black and white because it uses electrons instead of photons to view the specimen, and only light carries colour information. But black and white is not the truth, and I'm a perfectionist, so I try to mimic the natural colours. As a scientist, I think black and white hides some information, and as an artist, I want the image to be attractive. I try to highlight morphologically different structures to make them more visible so that the viewer can recognize complexity. I go to great lengths to reproduce the original colour.

How long does it take to create a final product?

It might take 20–60 hours — it depends on how much detail and how many structures are in the picture. It takes less time to create an image of a single pollen grain than of 5,000 spores of fungi, because I have to colour everything by hand.

Are you still active in research?

Yes. I have a year-long postdoctoral fellowship at the University of Basel, and I spend 20% of my time on cancer research. But my photography work frequently exposes me to samples and research from different areas of science — this year, for example, I produced an image from the dissection of a 1907 painting. ■

INTERVIEW BY KAREN KAPLAN

Editage in Treves, Pennsylvania, has websites in English, Japanese, German, Portuguese, Korean, simplified Chinese (for mainland China) and traditional Chinese (for Taiwan). The Japanese version has a free 'chat' function for scientists to ask general questions about publication and authorship practices; this function may soon be expanded to other languages. For a 4,000-word manuscript, Editage's basic language editing starts at around \$200; its premium service costs about \$400 and takes the paper from start to finish. Editors ensure that the article complies with the target journal's author guidelines and provide help with cover letters and responses to reviewer comments. They also provide coaching on issues such as journal selection, authorship, disclosure of conflicts of interest and peer-review function in Western journals.

Edanz, an editing firm headquartered in Hong Kong, caters mainly for scientists from Asia, the Middle East and South America. It offers comprehensive editing; a web-based journal-selection tool that can use a paper's abstract to draw up a list of the top ten most appropriate target journals from a pool of more than 10,000; a mock peer-review service; and in-person workshops on how to structure manuscripts.

"We teach authors how to be persuasive and construct stories that meet reader expectations," says Benjamin Shaw, China and global director of Edanz in Beijing. The company charges a flat fee based on the length and difficulty of the paper; articles below 4,000 words usually cost \$200–250. For that price, they do the work necessary — including multiple revisions — to bring it up to the level for publication in an international journal, says Tom da Costa, managing director for the firm's Japanese operations, based in Fukuoka.

Shaw notes that high-quality editing companies should be willing to answer questions about pricing, about who is doing the editing and about security and privacy. Edanz publishes its editors' profiles online so that scientists know who will be previewing their research.

Regardless of whether they seek help from colleagues, university services, private consultants or large firms, ESL scientists have to consider how to 'sell' their research to editors and reviewers, says Rashmi Nemade, founder of BioMedText, a biomedical writing and editing firm in Columbus, Ohio. "It's got to be entertaining and compelling," she says. "This concept is sometimes lost on scientists in general, but particularly those from other cultures where work is all about merit." ■

Kendall Powell is a freelance writer based in Lafayette, Colorado.

TURNING POINT

Martin Oeggerli

Best known as a scientific photographer, Martin Oeggerli takes close-up images of microscopic creatures and structures that have been featured in scientific publications and art galleries worldwide. Although he is still a postdoc in cancer research at the University of Basel in Switzerland, his photography has become his passion and his main career track. Oeggerli received his third Best Scientific Image prize in March, in The EMBO Journal's annual cover contest, for capturing the most interesting microstructure.

How did you come to take up photography?

I drew a lot as a child — I especially liked snakes and lizards — and I always tried to draw as accurately and in as much detail as I could. When I started to study, however, I didn't have much time to draw, so in 2001, when I was 26, my father gave me my first camera, a digital Nikon Coolpix 995. In one month, I took 20,000 images of insects, snakes and lizards.

How did technology affect your photography?

That first camera had just 3.2 megapixels, but it had an outstanding macro function and excellent post-processing software. I looked at a close-up picture I had taken of a fly and thought it was spectacular. So I kept trying to get closer, and started to buy micro-adapters for the camera. Finally, I turned to scanning electron microscopy (SEM). These images are the best: there is huge depth of field, which is very appealing because it resembles the depth we see with our own vision. The tiniest things are magnified 500,000 times or more.

When did you first get public recognition?

In 2006, I entered an international photography contest in Germany and submitted an image of a pollen grain germinating, just at the point of releasing a pollen tube. The pollen grain hydrates as soon as it attaches to the female part of the flower, and can grow the tube in a couple of seconds. I had tried to distinguish the different structures and present them in a way that looked natural. I didn't expect an award, but I won third place in Best Research Image.

Has there been broad interest in your work?

Yes. I grew up in a small town and didn't go to many art galleries, but I was interested in books and magazines. I started to read *National Geographic* when I was 12 or 13. I always liked the pictures — it is regarded as one of the world's most important photographic magazines. So I couldn't believe it when *National Geographic* published my images in December 2009 as part



A. TRAINA

of a feature about pollen grains. I started scientific microscopy as a hobby and never expected it to go anywhere, but this was my international breakthrough. It confirmed to me that the work I do really is of interest.

What was your EMBO winning entry?

An image of a network of mosquito eggs. By trapping air, this mesh-like web repels water, allows the eggs to float and keeps the whole structure from sinking.

You tint your images. Why?

SEM is always black and white because it uses electrons instead of photons to view the specimen, and only light carries colour information. But black and white is not the truth, and I'm a perfectionist, so I try to mimic the natural colours. As a scientist, I think black and white hides some information, and as an artist, I want the image to be attractive. I try to highlight morphologically different structures to make them more visible so that the viewer can recognize complexity. I go to great lengths to reproduce the original colour.

How long does it take to create a final product?

It might take 20–60 hours — it depends on how much detail and how many structures are in the picture. It takes less time to create an image of a single pollen grain than of 5,000 spores of fungi, because I have to colour everything by hand.

Are you still active in research?

Yes. I have a year-long postdoctoral fellowship at the University of Basel, and I spend 20% of my time on cancer research. But my photography work frequently exposes me to samples and research from different areas of science — this year, for example, I produced an image from the dissection of a 1907 painting. ■

INTERVIEW BY KAREN KAPLAN

WAITING FOR CHRONOMATIC

There's always time.

BY JEFF HECHT

I had almost given up hope, when the Chronomatic technician arrived at my little furnished flat in the twenty-first century. Four days had passed since both my Chronomatic Communicator and the Transporter itself had failed.

That was not supposed to happen. "Nobody ever gets lost in time with Chronomatic," the reviews said. If a Chronomatic Transporter failed, the Communicator could call help. Both would never fail, the sales bot had assured me, and authorized Chronomatic service technicians were available any place and any time that I might visit.

Sure enough, the Chronomatic emergency help number was on a pad next to the chronologically correct black telephone in my flat. After several minutes figuring out how to manipulate its silly antique dial, I was speaking with a bot, but it was so stupid that all it could do was say that Chronomatic would dispatch the first available technician if I left my name and location. Calling again reached the same robot, which said exactly the same thing.

I could only wait, taking short walks to get food, and killing time reading Wikipedia and watching YouTube on a dusty iPad. I was watching a Comcast technician sleeping on a couch when the doorbell finally rang. The short dark woman standing outside wore a blue Chronomatic Transport shirt, and carried a large shiny toolbox.

She apologized for the delay. "I am the only tech on call-down now, and I had to drive all the way from Kansas City. But there's always time."

"There's always time," I recalled, was Chronomatic's slogan.

I told her what had happened, and she sat on the couch, setting her toolbox on the coffee table. I watched from the armchair as she tested and probed the Transporter and Communicator. They did nothing. She got more tools from her van, and tested and probed some more.

Finally she looked at me, shrugged and said: "They're in a death loop."

"Can you fix it?"

"Of course not. It's a death loop. They are sending signals that make each other freeze."

"I pushed the reset before I called," I said.

"That didn't do any good. It never does.



It's a twenty-fourth century mistake; we can't fix it here."

I said something very impolite.

"Don't panic. I have replacement units in the van. All I have to do is call headquarters on my Communicator so that they can key in your ID and authorization codes for the new units. It won't take long once I get through."

She returned with a shiny new Transporter and a jet black Communicator. Their cases opened to show the screens, and I watched her wait for the start-up display. When all was ready, she sat on the couch and opened a Communicator that looked older and more worn than mine.

"Get me technician authorization transfer," she told the Communicator.

➤ NATURE.COM
Follow Futures on
Facebook at:
go.nature.com/mtoodm

It beeped once, then beeped again. I watched and waited.

"Still seeking connection," the Communicator said. It repeated the message ten more times before she cancelled the call to try again.

"Priority request. Technician requesting transfer of authorization to replacements for death-loop units," the technician said. This time it beeped twice, then announced: "Making connection."

I relaxed in the armchair. After four days, I could wait a few more minutes.

Half an hour passed. The Chronomatic technician sighed. "It sometimes takes a while."

I dozed off, waking with a start when I heard her talking. "I asked for the technician transfer-of-authorization line," she said, rather loudly. "What do you mean you don't do that any more?" She listened for a moment. "Go ahead and transfer me, then," she said, shaking her head.

Realizing that I was awake, she turned to me. "Wrong office. They reorganized support again, and didn't bother to tell anybody down in the twenty-first century."

This time she waited longer. The sun set, and I went out and got a pizza. She slipped a headset on while we ate. I stuck the last slice into the little refrigerator, to have for breakfast. When I came back she was talking to someone.

"His Transporter and Communicator are locked in a death loop. We're in the twenty-first century, and I cannot repair it," she said, rolling her eyes at me. "So who should I talk to?"

When she was finished, she put her communicator on the couch, stretched and said she needed to use the bathroom.

"Another office handles death loops," she said when she returned, and recited a new request. The communicator beeped twice, then announced: "Connection established. Please wait for the next available assistant. Estimated hold time 4 hours."

She shrugged and stretched out on the couch.

"Why can't a time-travel company route calls through time so they arrive when somebody is there to answer them?" I asked.

"Because they're a corporation," she said, and closed her eyes.

It is now 48 hours and counting. A Chronomatic technician is sleeping on my couch. ■

Jeff Hecht is Boston correspondent for *New Scientist* and a contributing editor to *Laser Focus World*.

PALAEOANTHROPOLOGY

The ancestral dinner table

Fossils from a new South African site show that some human ancestors ate fruits and leaves, as do most primates today. The finding challenges ideas of why and how the human lineage split from the ancestors of extant apes. [SEE LETTER P.90](#)

MARGARET J. SCHOENINGER

In 1871, Charles Darwin proposed¹ that our earliest ancestors lived in Africa alongside the ancestors of today's gorillas and chimpanzees, and ate a diet of fruit, leaves, seeds and nuts, similar to that of these extant primates. More recently, however, an alternative hypothesis² has taken precedence — that the human lineage split from the apes in part as a result of our ancestors' ability to obtain foods in open habitats, such as grasslands and savannah woodlands, that emerged in Africa following climatic changes during the Late Miocene epoch approximately 7 million years ago². These foods included grasses, sedge plants, grass-eating insects and small animals. On page 90 of this issue, Henry *et al.*³ present evidence that our early relatives had a more diverse diet, and ate items such as fruits, leaves and bark. The findings will trigger a rethink of the selective pressures that resulted in the separation of the ape and human lineages, and the traits we now consider to be unique to each.

Henry and colleagues' data derive from the fossils of two *Australopithecus sediba* individuals from Malapa in South Africa, which date to approximately 2 million years ago. *Au. sediba* was a species of australopithecine hominin — hominins are an evolutionary grouping that includes only humans and human ancestral species, which branched from the apes 4 million to 6 million years ago. The term australopithecine refers informally to species in the *Australopithecus* and *Paranthropus* genera, although some early *Homo* species are also referred to as australopithecine-like by some researchers⁴.

To build a picture of the likely diet of these species, researchers consider other fossils at hominin fossil sites, the dentistry of the hominin fossil teeth, and the proportions of stable isotopes of carbon that the teeth contain. This last technique gives an indication of the types of plants eaten, because different photosynthetic pathways differentially incorporate the two stable isotopes of carbon from carbon dioxide. Plants that use the C_4 photosynthetic pathway include mostly tropical grasses and some sedges; C_4 foods also include animals feeding on these plants, which are typical of savannah grasslands

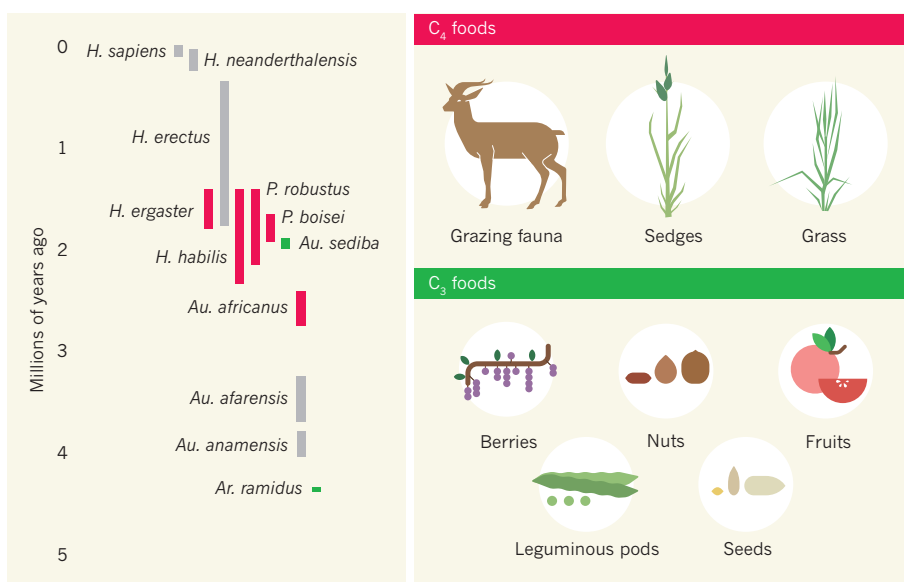


Figure 1 | Dietary detective work. The first species of hominin — extinct species belonging to the lineage that led to humans — are thought to have split from the lineage that led to extant primates approximately 4 million to 6 million years ago. Previous studies⁵ at African fossil sites had indicated that many hominins, including members of the *Homo*, *Paranthropus* and *Australopithecus* genera, primarily ate plants that use the C_4 photosynthetic pathway, and animals that fed on those plants. Such plants are characteristic of open habitats. Henry and colleagues³, however, describe the fossils of two *Australopithecus sediba* individuals that seem to have eaten a diet mostly based on C_3 plants and their products, as did the much earlier species *Ardipithecus ramidus*⁹. This diverse diet, which probably included nuts, berries, leaves and bark, suggests that *Au. sediba* lived in habitats that also contained trees and herbaceous plants, such as forests, woodlands and savannah woodlands.

and savannah woodlands. However, most plants, including trees and herbaceous plants that grow in forests, woodlands and savannah woodlands, use the C_3 photosynthetic pathway.

Previous studies of australopithecine fossil sites from eastern and southern Africa supported the view that *Australopithecus africanus*, *Paranthropus robustus*, *Paranthropus boisei*, *Homo habilis* and *Homo ergaster* all depended primarily on foods from habitats dissimilar from those of extant apes⁵. Today, chimpanzees live in forests, woodlands and savannah woodlands, although the latter are not favoured habitats⁶, and gorillas live in forests and dense woodlands. By contrast, studies at the fossil hominin sites implied that the hominins fed in open habitats and ate a wide range of C_4 foods, which made up approximately 35% of the diet of *Au. africanus*

and almost 100% of the foods consumed by the eastern African species *P. boisei*⁵ (Fig. 1).

This predominant C_4 signal led some researchers to suggest⁴ that “the consumption of C_4 foods is a fundamental hominin trait ... along with bipedalism”. But Henry and colleagues' study of *Au. sediba* suggests otherwise³. They propose that this species' feeding ecology was based on C_3 plants and included items such as fruits, herbaceous plants and bark (Fig. 1), and thereby differed in fundamental ways from the diet of all other australopithecines and later hominins studied so far.

The fossil fauna from the Malapa site includes grazing species, and the phytoliths (fossil plants) in the associated sediments are mostly C_4 species, pointing to the presence of grasslands in the region at the time. But the authors also found a fossilized fragment of

faeces, likely to be from a carnivore, which contains plant matter similar to that of modern trees that live along watercourses and in other humid regions, suggesting that this region also contained habitats with tree cover. In the *Au. sediba* fossils' dental plaque, the researchers identified 38 phytoliths, of which 15 seem to come from trees or herbaceous plants, 14 could not be identified and only 9 came from grasses. Carbon stable-isotope analysis of the teeth revealed a diet of exclusively, or almost exclusively, C_3 foods. Furthermore, the dental microwear patterns indicate that *Au. sediba* ingested hard objects, which the authors interpret as suggestive of feeding on tree bark, although nuts and fruits also seem likely possibilities.

These data suggest that *Au. sediba*'s dietary habits were similar to those of the majority of living primates, who eat predominantly C_3 foods⁷ and seek these out even in habitats that include significant regions of C_4 grasses^{7,8}. The findings also place *Au. sediba* as feeding in ways similar to *Ardipithecus ramidus*, an earlier hominin species from 4.4 million years ago that probably also fed in savannah woodland interspersed with forest patches and grassland⁹, just as has been hypothesized for the first hominins⁶.

The significance of these results extends beyond whether a diet based on C_4 foods is a fundamental hominin trait. It also brings into question our understanding of the evolution of bipedalism — another trait that is thought of as being fundamentally human. The species in which bipedalism emerged, and the evolutionary pressures that drove this adaptation, remain topics of debate. So far, it seems that *Ar. ramidus* used some form of bipedalism⁹; the 3.95-million-year-old *Australopithecus anamensis* may have walked upright when necessary³; the 3.8-million-year-old *Australopithecus afarensis* could walk bipedally for limited distances³; and a series of hominins dating to approximately 2 million years ago (*P. boisei*, *H. habilis*, *H. ergaster* and *Homo rudolfensis*) were all bipedal to varying extents³. But exactly where these species fit on genetic-relatedness trees is not clear. *Au. sediba* was bipedal, but in a way that differs from hominins that inhabited eastern Africa, and the species retained some ability to climb¹⁰.

How does the suggestion of dietary variation among hominins fit with this understanding of bipedalism and the branching of hominin species? Researchers have suggested that there was an adaptive radiation event approximately 2 million years ago¹⁰, in which a common ancestor relatively rapidly gave rise to a range of bipedal species with novel traits that allowed them to move into new habitats. The unexamined assumption that all bipedal species share a unique genealogical relationship is reminiscent of evolutionary biologist Stephen J. Gould's question¹¹: "What, if anything, is a zebra?" Gould examined whether the three species

of modern zebra are more closely related to one another than to other non-striped horse species. He concluded that striping evolved only once. We should also ask: what, if anything, is a hominin? Perhaps all ancestral ape-like species that walked on two legs and had a C_4 -focused diet were uniquely related to each other but were not necessarily human ancestors. Maybe humans emerged from some other hominin groups around 2 million years ago that were also bipedal and had more general, opportunistic foraging strategies, including meat-eating. Only enterprising studies into other aspects of australopithecine and hominin life, like Henry and colleagues' analysis of *Au. sediba*'s diet, will provide us with definitive answers. ■

Margaret J. Schoeninger is in the Department of Anthropology,

University of California, San Diego, La Jolla, California 92093, USA. e-mail: mjschoen@ucsd.edu

1. Darwin, C. *The Descent of Man, and Selection in Relation to Sex* (Murray, 1871).
2. Cerling, T. E. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **97**, 241–247 (1992).
3. Henry, A. G. et al. *Nature* **487**, 90–93 (2012).
4. Wood, B. & Leakey, M. *Evol. Anthropol.* **20**, 264–292 (2011).
5. Sponheimer, M. et al. *J. Hum. Evol.* **48**, 301–312 (2005).
6. Moore, J. in *Topics in Primatology: Human Origins* (eds Nishida, T. et al.) 99–118 (Univ. Tokyo Press, 1992).
7. Codron, D., Lee-Thorp, J. A., Sponheimer, M., de Ruiter, D. & Codron, J. *Am. J. Phys. Anthropol.* **129**, 204–214 (2006).
8. Schoeninger, M. J., Moore, J. & Sept, J. M. *Am. J. Primatol.* **49**, 297–314 (1999).
9. White, T. D. et al. *Science* **326**, 75–86 (2009).
10. Pickering, R. et al. *Science* **333**, 1421–1423 (2011).
11. Gould, S. J. in *Hen's Teeth And Horse's Toes* Ch. 28, 355–365 (Norton, 1983).

STEM CELLS

A sporadic super state

It seems that embryonic stem cells regularly pass through a transient state during which they can generate all the cell types of an animal, including those of the placenta. [SEE ARTICLE P.57](#)

AZIM SURANI & JULIA TISCHLER

During mammalian development, the fertilized egg (zygote) and its descendant cells up to the eight-cell stage are totipotent — any single cell can differentiate to form all the cells of an organism, and give rise to fetal and placental tissues. This property, however, is lost by the time development has reached the 32-cell blastocyst stage. Therefore, embryonic stem (ES) cells, which are derived from the inner cell mass of blastocysts, are not considered to be totipotent, but pluripotent: they can produce all the cell types of a fetus, but rarely those of the placenta. However, Macfarlan *et al.*¹ (page 57) show that some cells in ES-cell cultures can differentiate into both embryonic and extraembryonic tissues. What's more, most, if not all, ES cells regularly transit this 'totipotent-like' state.

Maternally inherited gene products initially control cellular processes in mammalian zygotes, and it is only at the two-cell stage that the zygote's own genome is activated and assumes full control of development. Interestingly, many of the genes expressed at this stage are retroelements² — DNA sequences of viral origin that can duplicate themselves and 'jump' into other locations in the genome. Indeed, Macfarlan *et al.* found that expression of many of the genes activated at the two-cell stage in mouse embryos is driven by regulatory sequences derived from a type of retroelement known as MERVL. Such genes

determine the fate of the precursors of placental and fetal cells shortly thereafter (Fig. 1a).

To label the two-cell stage, the authors introduced into zygotes a synthetic 'reporter' gene that contained a MERVL regulatory sequence and also encoded a red fluorescent protein. As expected, the reporter's expression was intense at the two-cell stage but then gradually faded away and was not observed in blastocysts. However, cultures of ES cells into which the reporter gene had been inserted unexpectedly also contained a few cells (about 0.2–1.5%) that produced the fluorescent protein. Moreover, these rare cells had an overall gene-expression profile that was similar to that of two-cell embryos and was clearly different from that of the rest of the ES cells in the cultures.

Macfarlan *et al.* found the unusual 'two-cell-like' ES cells under all the culture conditions they tested, although in variable numbers. Most importantly, they showed that most, if not all, ES cells passed through a transient two-cell-like state at some time (Fig. 1b). This result is consistent with previous findings of fluctuating expression of other reporter genes in ES cells^{3–8}. Notably, the authors observed cells with similar properties in cultures of induced pluripotent stem cells — such cells are not obtained from blastocysts but by 'reprogramming' differentiated cells such as fibroblasts. Therefore, the presence in ES-cell cultures of a few cells showing two-cell-like gene expression was not due to contamination with cells from early-stage embryos, but is a feature that is probably

faeces, likely to be from a carnivore, which contains plant matter similar to that of modern trees that live along watercourses and in other humid regions, suggesting that this region also contained habitats with tree cover. In the *Au. sediba* fossils' dental plaque, the researchers identified 38 phytoliths, of which 15 seem to come from trees or herbaceous plants, 14 could not be identified and only 9 came from grasses. Carbon stable-isotope analysis of the teeth revealed a diet of exclusively, or almost exclusively, C_3 foods. Furthermore, the dental microwear patterns indicate that *Au. sediba* ingested hard objects, which the authors interpret as suggestive of feeding on tree bark, although nuts and fruits also seem likely possibilities.

These data suggest that *Au. sediba*'s dietary habits were similar to those of the majority of living primates, who eat predominantly C_3 foods⁷ and seek these out even in habitats that include significant regions of C_4 grasses^{7,8}. The findings also place *Au. sediba* as feeding in ways similar to *Ardipithecus ramidus*, an earlier hominin species from 4.4 million years ago that probably also fed in savannah woodland interspersed with forest patches and grassland⁹, just as has been hypothesized for the first hominins⁶.

The significance of these results extends beyond whether a diet based on C_4 foods is a fundamental hominin trait. It also brings into question our understanding of the evolution of bipedalism — another trait that is thought of as being fundamentally human. The species in which bipedalism emerged, and the evolutionary pressures that drove this adaptation, remain topics of debate. So far, it seems that *Ar. ramidus* used some form of bipedalism⁹; the 3.95-million-year-old *Australopithecus anamensis* may have walked upright when necessary³; the 3.8-million-year-old *Australopithecus afarensis* could walk bipedally for limited distances³; and a series of hominins dating to approximately 2 million years ago (*P. boisei*, *H. habilis*, *H. ergaster* and *Homo rudolfensis*) were all bipedal to varying extents³. But exactly where these species fit on genetic-relatedness trees is not clear. *Au. sediba* was bipedal, but in a way that differs from hominins that inhabited eastern Africa, and the species retained some ability to climb¹⁰.

How does the suggestion of dietary variation among hominins fit with this understanding of bipedalism and the branching of hominin species? Researchers have suggested that there was an adaptive radiation event approximately 2 million years ago¹⁰, in which a common ancestor relatively rapidly gave rise to a range of bipedal species with novel traits that allowed them to move into new habitats. The unexamined assumption that all bipedal species share a unique genealogical relationship is reminiscent of evolutionary biologist Stephen J. Gould's question¹¹: "What, if anything, is a zebra?" Gould examined whether the three species

of modern zebra are more closely related to one another than to other non-striped horse species. He concluded that striping evolved only once. We should also ask: what, if anything, is a hominin? Perhaps all ancestral ape-like species that walked on two legs and had a C_4 -focused diet were uniquely related to each other but were not necessarily human ancestors. Maybe humans emerged from some other hominin groups around 2 million years ago that were also bipedal and had more general, opportunistic foraging strategies, including meat-eating. Only enterprising studies into other aspects of australopithecine and hominin life, like Henry and colleagues' analysis of *Au. sediba*'s diet, will provide us with definitive answers. ■

Margaret J. Schoeninger is in the Department of Anthropology,

University of California, San Diego, La Jolla, California 92093, USA. e-mail: mjschoen@ucsd.edu

1. Darwin, C. *The Descent of Man, and Selection in Relation to Sex* (Murray, 1871).
2. Cerling, T. E. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **97**, 241–247 (1992).
3. Henry, A. G. et al. *Nature* **487**, 90–93 (2012).
4. Wood, B. & Leakey, M. *Evol. Anthropol.* **20**, 264–292 (2011).
5. Sponheimer, M. et al. *J. Hum. Evol.* **48**, 301–312 (2005).
6. Moore, J. in *Topics in Primatology: Human Origins* (eds Nishida, T. et al.) 99–118 (Univ. Tokyo Press, 1992).
7. Codron, D., Lee-Thorp, J. A., Sponheimer, M., de Ruiter, D. & Codron, J. *Am. J. Phys. Anthropol.* **129**, 204–214 (2006).
8. Schoeninger, M. J., Moore, J. & Sept, J. M. *Am. J. Primatol.* **49**, 297–314 (1999).
9. White, T. D. et al. *Science* **326**, 75–86 (2009).
10. Pickering, R. et al. *Science* **333**, 1421–1423 (2011).
11. Gould, S. J. in *Hen's Teeth And Horse's Toes* Ch. 28, 355–365 (Norton, 1983).

STEM CELLS

A sporadic super state

It seems that embryonic stem cells regularly pass through a transient state during which they can generate all the cell types of an animal, including those of the placenta. [SEE ARTICLE P.57](#)

AZIM SURANI & JULIA TISCHLER

During mammalian development, the fertilized egg (zygote) and its descendant cells up to the eight-cell stage are totipotent — any single cell can differentiate to form all the cells of an organism, and give rise to fetal and placental tissues. This property, however, is lost by the time development has reached the 32-cell blastocyst stage. Therefore, embryonic stem (ES) cells, which are derived from the inner cell mass of blastocysts, are not considered to be totipotent, but pluripotent: they can produce all the cell types of a fetus, but rarely those of the placenta. However, Macfarlan *et al.*¹ (page 57) show that some cells in ES-cell cultures can differentiate into both embryonic and extraembryonic tissues. What's more, most, if not all, ES cells regularly transit this 'totipotent-like' state.

Maternally inherited gene products initially control cellular processes in mammalian zygotes, and it is only at the two-cell stage that the zygote's own genome is activated and assumes full control of development. Interestingly, many of the genes expressed at this stage are retroelements² — DNA sequences of viral origin that can duplicate themselves and 'jump' into other locations in the genome. Indeed, Macfarlan *et al.* found that expression of many of the genes activated at the two-cell stage in mouse embryos is driven by regulatory sequences derived from a type of retroelement known as MERVL. Such genes

determine the fate of the precursors of placental and fetal cells shortly thereafter (Fig. 1a).

To label the two-cell stage, the authors introduced into zygotes a synthetic 'reporter' gene that contained a MERVL regulatory sequence and also encoded a red fluorescent protein. As expected, the reporter's expression was intense at the two-cell stage but then gradually faded away and was not observed in blastocysts. However, cultures of ES cells into which the reporter gene had been inserted unexpectedly also contained a few cells (about 0.2–1.5%) that produced the fluorescent protein. Moreover, these rare cells had an overall gene-expression profile that was similar to that of two-cell embryos and was clearly different from that of the rest of the ES cells in the cultures.

Macfarlan *et al.* found the unusual 'two-cell-like' ES cells under all the culture conditions they tested, although in variable numbers. Most importantly, they showed that most, if not all, ES cells passed through a transient two-cell-like state at some time (Fig. 1b). This result is consistent with previous findings of fluctuating expression of other reporter genes in ES cells^{3–8}. Notably, the authors observed cells with similar properties in cultures of induced pluripotent stem cells — such cells are not obtained from blastocysts but by 'reprogramming' differentiated cells such as fibroblasts. Therefore, the presence in ES-cell cultures of a few cells showing two-cell-like gene expression was not due to contamination with cells from early-stage embryos, but is a feature that is probably



50 Years Ago

To many people nutrition means no more than the study of foods and their use by the living body. But with increasing evidence of widespread malnutrition throughout the world, and of a deteriorating situation in regard to world population and food supplies, much emphasis needs to be placed on the importance of work outside the laboratory on problems associated with providing people everywhere with an adequate diet. In many parts of the world an immediate problem is a means of conveying foods from one place to another. For example, fruit, the sale of which provides a livelihood for some people, may be rotting on the trees because there are no roads on which it can be transported to other people who might benefit by addition of fruit to their diet ... Lack of adequate means of distribution may seriously limit the use of fish, and while many people suffer the effects of protein malnutrition, fishermen not far away suffer economic disaster because they cannot get rid of harvest gluts.

From *Nature* 7 July 1962

100 Years Ago

Measurements of the temperature of flowing lava are so rare that some made by Prof. G. Platania during the eruption of Etna last September possess considerable interest ... His observations were made with a F ry's radio-pyrometer on a stream of lava flowing from the lowest of a string of craters in the neighbourhood of M. Rosso, a few days before the eruption ceased. The temperatures, in parts where the lava was still red, ranged from 795  to a maximum of 940  C.

From *Nature* 4 July 1912

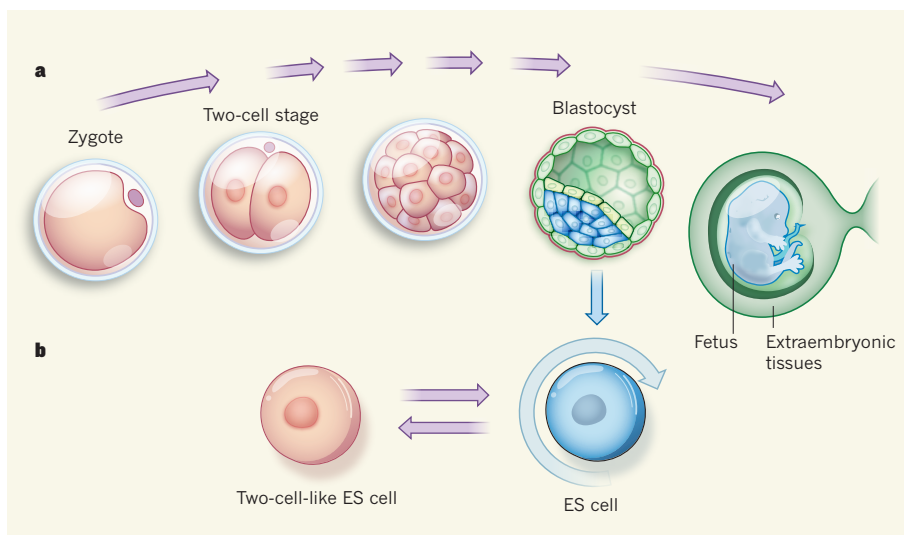


Figure 1 | A 'totipotent-like' state in embryonic stem cells. **a**, During the early development of a mouse embryo, many genes are transiently expressed after the first division of the fertilized egg (zygote). Macfarlan *et al.*¹ find that some of those genes contain virus-derived sequences that drive the genes' transient expression at the two-cell stage. Further cellular divisions result in the development of a blastocyst, a mass of cells that have defined fates: those in the inner cell mass give rise to the fetus, whereas the rest contribute to extraembryonic tissues such as the placenta. **b**, Embryonic stem (ES) cells, which are derived from the inner cell mass of blastocysts, self-renew in culture (curved arrow). The authors show that most, if not all, ES cells pass through a short-lived state during which they display features that are typical of the totipotent two-cell stage: unlike the rest of the ES cells in the culture, they lack expression of the proteins Oct4, Sox2 and Nanog, and have the ability to form cells of both the placenta and the fetus.

shared by all pluripotent stem cells.

Crucially, the researchers established that the rare ES cells, but not the rest of the ES cells in a culture, can contribute to both fetal and placental tissues, thereby fulfilling a key attribute of totipotency. However, it remains to be seen whether the two-cell-like ES cells have the potential to generate a complete, live organism.

What is the significance of the two-cell-like ES cells? To address this question, it is important to note that cells from two-cell embryos or early blastocysts represent transient states. Unlike ES cells, they do not self-renew but progress to the next developmental stage. Furthermore, both symmetrical and asymmetrical cell divisions follow the two-cell stage and establish which blastocyst cells will develop into the fetus and which will become extraembryonic tissues⁹. By contrast, ES cells self-renew through symmetrical divisions. So how can the two-cell-like ES cells generate extraembryonic tissues? A precise analysis of single ES cells is required to understand how they generate different cell types, and whether the cells' fate is determined stochastically or is a pre-programmed property of individual cells. For example, do key molecular determinants of different fetal and placental lineages become segregated into individual cells when they differentiate from two-cell-like ES cells?

The authors found that the two-cell-like cells, unlike the rest of the ES cells, did not produce the proteins Oct4, Sox2 and Nanog, which are typically associated with pluripotency, and instead expressed several genes that are commonly active in two-cell embryos.

Notably, one of these genes encodes the protein Zscan4, which is required for the maintenance of telomeres (the ends of chromosomes, which are eroded every time DNA duplicates) and for genomic stability. Lack of Zscan4 leads to a gradual decline in the proliferative capacity of ES cells¹⁰. Macfarlan *et al.* observed that the rare ES cells also displayed a two-cell-like pattern of epigenomic marks — chemical modifications of DNA, and of DNA-bound proteins, that do not alter DNA sequence but affect gene expression.

Therefore, fluctuating patterns of gene expression might provide pluripotent stem cells with a window of opportunity to enter specific cell states. In particular, transition of ES cells through a two-cell-like state may be crucial for 'resetting' the epigenome, for the repair and maintenance of telomeres, and for refreshing the core genetic network underlying pluripotency. Future research, particularly at the single-cell level, may help to reveal why, and how, these cells go through such fluctuating states. It may also advance our knowledge of the mechanisms of cellular rejuvenation and reprogramming in early germ cells, which eventually develop into sperm and eggs. These cells are considered to be immortal, as they have the potential to generate a whole organism and therefore all subsequent generations. ■

Azim Surani and Julia Tischler are at the Wellcome Trust Cancer Research UK Gurdon Institute, University of Cambridge, Cambridge CB2 1QN, UK.
e-mail: a.surani@gurdon.cam.ac.uk

- Macfarlan, T. S. *et al. Nature* **487**, 57–63 (2012).
- Peaston, A. E. *et al. Dev. Cell* **7**, 597–606 (2004).
- Toyooka, Y., Shimosato, D., Murakami, K., Takahashi, K. & Niwa, H. *Development* **135**, 909–918 (2008).
- Miyazaki, Y. & Torres-Padilla, M.-E. *Nature* **483**, 470–473 (2012).
- Kalmar, T. *et al. PLoS Biol.* **7**, e1000149 (2009).
- Hayashi, K., Chuva de Sousa Lopes, S. M., Tang, F. & Surani, M. A. *Cell Stem Cell* **3**, 391–401 (2008).
- Chambers, I. *et al. Nature* **450**, 1230–1234 (2007).
- Canham, M. A., Sharov, A. A., Ko, M. S. H. & Brickman, J. M. *PLoS Biol.* **8**, e1000379 (2010).
- Rossant, J. & Tam, P. P. L. *Development* **136**, 701–713 (2009).
- Zalzman, M. *et al. Nature* **464**, 858–863 (2010).

ASTRONOMY

Warm dust makes a fast getaway

A rapid drop in infrared emission from a Sun-like star could indicate that a drastic event has cleared a circumstellar disk of dusty debris — the material from which planets form. [SEE LETTER P.74](#)

MARGARET MOERCHEN

The formation and destruction of planetary bodies such as asteroids and comets are the main events that sculpt the dusty disks around young stars. During these processes, large quantities of dust are swept from the disk into nascent planets (planetesimals), paths are cleared by these bodies' orbits, and material is re-released from existing planetesimals through destructive collisions or sublimation. We know that such processes were involved in forming the architecture of the Solar System, as well as that of the ever-increasing number of planetary systems being discovered around stars other than the Sun¹. But the precise timescales and conditions required for the formation of planets in the disks are still under investigation. On page 74 of this issue, Melis *et al.*² report that the infrared radiation attributed to a dusty disk around a Sun-like star has diminished by a factor of about 30 in less than 2 years. The researchers propose that this reduction has arisen because of a correspondingly drastic depletion of the dust disk through the mechanisms outlined above. Although the data are not yet supported by any theoretical model, the system in question is remarkable for the speed with which its surrounding material seems to have disappeared.

Circumstellar dust disks are revealed by the presence of infrared emission in excess of that expected from the star alone: the excess emission is a signature of dust that has been heated by the star³. Evidence for planet-forming activity in the disk is charted in the rise and fall of the disk's luminosity, which typically

increases for the first few million years of the star's lifetime, begins to peak when the star is around 10 million years old, and then steadily declines⁴. The Sun-like star observed by Melis *et al.*², known as TYC 8241 2652 1, is 10 million years old. It would therefore be expected to host an especially active disk and to offer an interesting environment in which to study physical processes.

It is generally accepted that small planetesimals of rock and ice can grow to a critical size that is capable of gravitationally stirring the disk and thereby increasing the mostly collisional activity that generates more dust. Comet-like bodies of dust and ice residing in the disk can also be perturbed by the

gravitational influence of larger planetary bodies and can end up near the star, resulting in sublimation of the smaller bodies' volatile components and the release of dust. This dust is eventually ground down to particles small enough to be radiatively blown out of the system and, as a result, the disk's brightness decays. Notably, even for some disks in which the amount of dust present is considered likely to be transient and evolving rapidly, significant changes in that amount are expected to take at least thousands of years⁵.

By comparing visible-light data from the Tycho-2 catalogue⁶ with infrared data collected by the IRAS (ref. 7) and AKARI (ref. 8) missions in 1983 and 2006, respectively, Melis *et al.* found that 11% of the total light emitted by TYC 8241 2652 1 is intercepted by dust and re-radiated in the infrared. This infrared radiation, which was detected before the recently observed decay in infrared brightness, is best fitted by a model describing thermal emission from warm material at a temperature of about 450 kelvin (see Fig. 1 of the paper²) and at a distance from the star of roughly 40% of the average distance between Earth and the Sun.

Follow-up observations — obtained by Melis *et al.* in 2008 using the Thermal-Region Camera Spectrograph⁹ (T-ReCS) at the 8-metre Gemini South telescope in Chile — initially revealed disk-brightness levels of about 300–400 millijanskys, similar to those recorded by the satellites IRAS and AKARI. These higher-resolution ground-based observations rule out the possibility that the infrared radiation originated not from the disk but from a source in the same line of sight as the satellites. Measurements made at Gemini eight months later yielded only one-third of the original disk brightness, and this level decreased a further tenfold (to less than 20 millijanskys) in images from the Wide-field Infrared Space Explorer¹⁰ (WISE) in 2010. This level was confirmed at Gemini as recently as May 2012 (Fig. 1).

A sharp drop in luminosity on a timescale of years might signal variability due to periodic accretion of gas onto the star or to intra-disk shadowing, in which an inner region of the disk puffs up through stellar heating and prevents light from reaching the more distant parts of the disk. However, these hypotheses (along with the possibility of occulting material lying somewhere along our line of sight to the star) can be excluded on the basis of the stability of the star's visible and near-infrared emission over the period considered. The authors instead argue in favour of two models involving previously established processes of disk and planetary evolution to explain the luminosity drop and the apparent disappearance of dust.

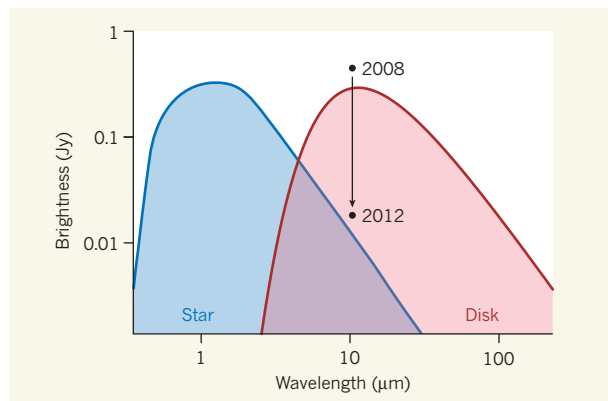


Figure 1 | Dimming disk. In this plot of brightness (flux density, given in janskys) against wavelength, the two shaded areas represent the approximate emission from the Sun-like star studied by Melis *et al.*² and the dust disk that surrounds it. The dust is heated by the star, which has a temperature of about 5,000 kelvin, but the dust is nonetheless much cooler, at roughly 450 kelvin. The peak emission of the dust occurs at a longer wavelength — in the mid-infrared — than that of the star. Two measurements of the dust brightness at about 10 micrometres, from the same telescope, are shown here: one was obtained in 2008 and the other in 2012. The brightness measured in 2012 was nearly 30 times smaller than that in 2008. This rapid decline in brightness is attributed to an unusually fast removal of dust from the system.



50 Years Ago

To many people nutrition means no more than the study of foods and their use by the living body. But with increasing evidence of widespread malnutrition throughout the world, and of a deteriorating situation in regard to world population and food supplies, much emphasis needs to be placed on the importance of work outside the laboratory on problems associated with providing people everywhere with an adequate diet. In many parts of the world an immediate problem is a means of conveying foods from one place to another. For example, fruit, the sale of which provides a livelihood for some people, may be rotting on the trees because there are no roads on which it can be transported to other people who might benefit by addition of fruit to their diet ... Lack of adequate means of distribution may seriously limit the use of fish, and while many people suffer the effects of protein malnutrition, fishermen not far away suffer economic disaster because they cannot get rid of harvest gluts.

From *Nature* 7 July 1962

100 Years Ago

Measurements of the temperature of flowing lava are so rare that some made by Prof. G. Platania during the eruption of Etna last September possess considerable interest ... His observations were made with a F ry's radio-pyrometer on a stream of lava flowing from the lowest of a string of craters in the neighbourhood of M. Rosso, a few days before the eruption ceased. The temperatures, in parts where the lava was still red, ranged from 795  to a maximum of 940  C.

From *Nature* 4 July 1912

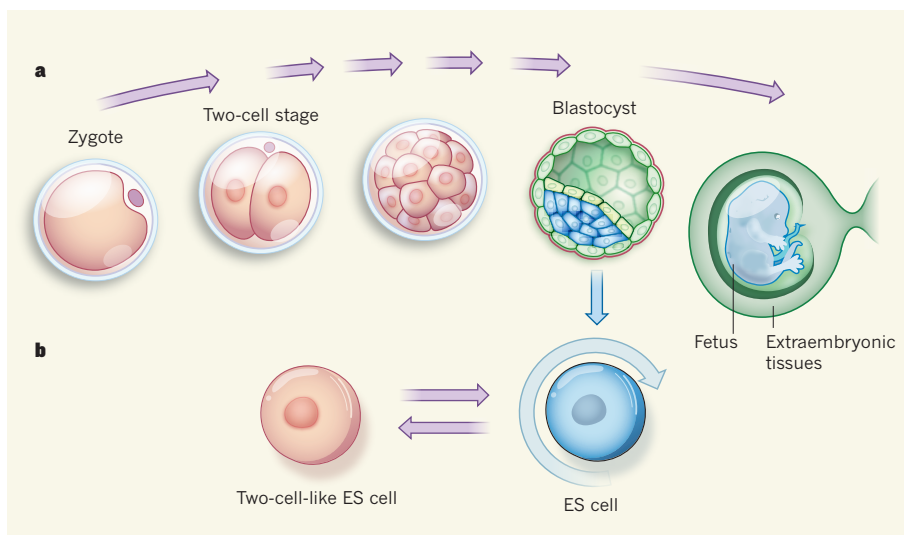


Figure 1 | A 'totipotent-like' state in embryonic stem cells. **a**, During the early development of a mouse embryo, many genes are transiently expressed after the first division of the fertilized egg (zygote). Macfarlan *et al.*¹ find that some of those genes contain virus-derived sequences that drive the genes' transient expression at the two-cell stage. Further cellular divisions result in the development of a blastocyst, a mass of cells that have defined fates: those in the inner cell mass give rise to the fetus, whereas the rest contribute to extraembryonic tissues such as the placenta. **b**, Embryonic stem (ES) cells, which are derived from the inner cell mass of blastocysts, self-renew in culture (curved arrow). The authors show that most, if not all, ES cells pass through a short-lived state during which they display features that are typical of the totipotent two-cell stage: unlike the rest of the ES cells in the culture, they lack expression of the proteins Oct4, Sox2 and Nanog, and have the ability to form cells of both the placenta and the fetus.

shared by all pluripotent stem cells.

Crucially, the researchers established that the rare ES cells, but not the rest of the ES cells in a culture, can contribute to both fetal and placental tissues, thereby fulfilling a key attribute of totipotency. However, it remains to be seen whether the two-cell-like ES cells have the potential to generate a complete, live organism.

What is the significance of the two-cell-like ES cells? To address this question, it is important to note that cells from two-cell embryos or early blastocysts represent transient states. Unlike ES cells, they do not self-renew but progress to the next developmental stage. Furthermore, both symmetrical and asymmetrical cell divisions follow the two-cell stage and establish which blastocyst cells will develop into the fetus and which will become extraembryonic tissues⁹. By contrast, ES cells self-renew through symmetrical divisions. So how can the two-cell-like ES cells generate extraembryonic tissues? A precise analysis of single ES cells is required to understand how they generate different cell types, and whether the cells' fate is determined stochastically or is a pre-programmed property of individual cells. For example, do key molecular determinants of different fetal and placental lineages become segregated into individual cells when they differentiate from two-cell-like ES cells?

The authors found that the two-cell-like cells, unlike the rest of the ES cells, did not produce the proteins Oct4, Sox2 and Nanog, which are typically associated with pluripotency, and instead expressed several genes that are commonly active in two-cell embryos.

Notably, one of these genes encodes the protein Zscan4, which is required for the maintenance of telomeres (the ends of chromosomes, which are eroded every time DNA duplicates) and for genomic stability. Lack of Zscan4 leads to a gradual decline in the proliferative capacity of ES cells¹⁰. Macfarlan *et al.* observed that the rare ES cells also displayed a two-cell-like pattern of epigenomic marks — chemical modifications of DNA, and of DNA-bound proteins, that do not alter DNA sequence but affect gene expression.

Therefore, fluctuating patterns of gene expression might provide pluripotent stem cells with a window of opportunity to enter specific cell states. In particular, transition of ES cells through a two-cell-like state may be crucial for 'resetting' the epigenome, for the repair and maintenance of telomeres, and for refreshing the core genetic network underlying pluripotency. Future research, particularly at the single-cell level, may help to reveal why, and how, these cells go through such fluctuating states. It may also advance our knowledge of the mechanisms of cellular rejuvenation and reprogramming in early germ cells, which eventually develop into sperm and eggs. These cells are considered to be immortal, as they have the potential to generate a whole organism and therefore all subsequent generations. ■

Azim Surani and Julia Tischler are at the Wellcome Trust Cancer Research UK Gurdon Institute, University of Cambridge, Cambridge CB2 1QN, UK.
e-mail: a.surani@gurdon.cam.ac.uk

- Macfarlan, T. S. *et al. Nature* **487**, 57–63 (2012).
- Peaston, A. E. *et al. Dev. Cell* **7**, 597–606 (2004).
- Toyooka, Y., Shimosato, D., Murakami, K., Takahashi, K. & Niwa, H. *Development* **135**, 909–918 (2008).
- Miyazaki, Y. & Torres-Padilla, M.-E. *Nature* **483**, 470–473 (2012).
- Kalmar, T. *et al. PLoS Biol.* **7**, e1000149 (2009).
- Hayashi, K., Chuva de Sousa Lopes, S. M., Tang, F. & Surani, M. A. *Cell Stem Cell* **3**, 391–401 (2008).
- Chambers, I. *et al. Nature* **450**, 1230–1234 (2007).
- Canham, M. A., Sharov, A. A., Ko, M. S. H. & Brickman, J. M. *PLoS Biol.* **8**, e1000379 (2010).
- Rossant, J. & Tam, P. P. L. *Development* **136**, 701–713 (2009).
- Zalzman, M. *et al. Nature* **464**, 858–863 (2010).

ASTRONOMY

Warm dust makes a fast getaway

A rapid drop in infrared emission from a Sun-like star could indicate that a drastic event has cleared a circumstellar disk of dusty debris — the material from which planets form. SEE LETTER P.74

MARGARET MOERCHEN

The formation and destruction of planetary bodies such as asteroids and comets are the main events that sculpt the dusty disks around young stars. During these processes, large quantities of dust are swept from the disk into nascent planets (planetesimals), paths are cleared by these bodies' orbits, and material is re-released from existing planetesimals through destructive collisions or sublimation. We know that such processes were involved in forming the architecture of the Solar System, as well as that of the ever-increasing number of planetary systems being discovered around stars other than the Sun¹. But the precise timescales and conditions required for the formation of planets in the disks are still under investigation. On page 74 of this issue, Melis *et al.*² report that the infrared radiation attributed to a dusty disk around a Sun-like star has diminished by a factor of about 30 in less than 2 years. The researchers propose that this reduction has arisen because of a correspondingly drastic depletion of the dust disk through the mechanisms outlined above. Although the data are not yet supported by any theoretical model, the system in question is remarkable for the speed with which its surrounding material seems to have disappeared.

Circumstellar dust disks are revealed by the presence of infrared emission in excess of that expected from the star alone: the excess emission is a signature of dust that has been heated by the star³. Evidence for planet-forming activity in the disk is charted in the rise and fall of the disk's luminosity, which typically

increases for the first few million years of the star's lifetime, begins to peak when the star is around 10 million years old, and then steadily declines⁴. The Sun-like star observed by Melis *et al.*², known as TYC 8241 2652 1, is 10 million years old. It would therefore be expected to host an especially active disk and to offer an interesting environment in which to study physical processes.

It is generally accepted that small planetesimals of rock and ice can grow to a critical size that is capable of gravitationally stirring the disk and thereby increasing the mostly collisional activity that generates more dust. Comet-like bodies of dust and ice residing in the disk can also be perturbed by the

gravitational influence of larger planetary bodies and can end up near the star, resulting in sublimation of the smaller bodies' volatile components and the release of dust. This dust is eventually ground down to particles small enough to be radiatively blown out of the system and, as a result, the disk's brightness decays. Notably, even for some disks in which the amount of dust present is considered likely to be transient and evolving rapidly, significant changes in that amount are expected to take at least thousands of years⁵.

By comparing visible-light data from the Tycho-2 catalogue⁶ with infrared data collected by the IRAS (ref. 7) and AKARI (ref. 8) missions in 1983 and 2006, respectively, Melis *et al.* found that 11% of the total light emitted by TYC 8241 2652 1 is intercepted by dust and re-radiated in the infrared. This infrared radiation, which was detected before the recently observed decay in infrared brightness, is best fitted by a model describing thermal emission from warm material at a temperature of about 450 kelvin (see Fig. 1 of the paper²) and at a distance from the star of roughly 40% of the average distance between Earth and the Sun.

Follow-up observations — obtained by Melis *et al.* in 2008 using the Thermal-Region Camera Spectrograph⁹ (T-ReCS) at the 8-metre Gemini South telescope in Chile — initially revealed disk-brightness levels of about 300–400 millijanskys, similar to those recorded by the satellites IRAS and AKARI. These higher-resolution ground-based observations rule out the possibility that the infrared radiation originated not from the disk but from a source in the same line of sight as the satellites. Measurements made at Gemini eight months later yielded only one-third of the original disk brightness, and this level decreased a further tenfold (to less than 20 millijanskys) in images from the Wide-field Infrared Space Explorer¹⁰ (WISE) in 2010. This level was confirmed at Gemini as recently as May 2012 (Fig. 1).

A sharp drop in luminosity on a timescale of years might signal variability due to periodic accretion of gas onto the star or to intra-disk shadowing, in which an inner region of the disk puffs up through stellar heating and prevents light from reaching the more distant parts of the disk. However, these hypotheses (along with the possibility of occulting material lying somewhere along our line of sight to the star) can be excluded on the basis of the stability of the star's visible and near-infrared emission over the period considered. The authors instead argue in favour of two models involving previously established processes of disk and planetary evolution to explain the luminosity drop and the apparent disappearance of dust.

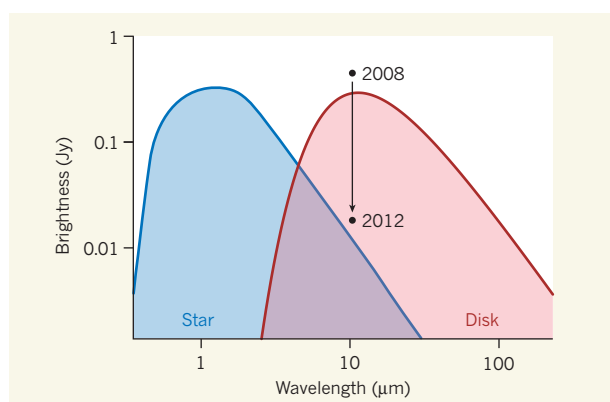


Figure 1 | Dimming disk. In this plot of brightness (flux density, given in janskys) against wavelength, the two shaded areas represent the approximate emission from the Sun-like star studied by Melis *et al.*² and the dust disk that surrounds it. The dust is heated by the star, which has a temperature of about 5,000 kelvin, but the dust is nonetheless much cooler, at roughly 450 kelvin. The peak emission of the dust occurs at a longer wavelength — in the mid-infrared — than that of the star. Two measurements of the dust brightness at about 10 micrometres, from the same telescope, are shown here: one was obtained in 2008 and the other in 2012. The brightness measured in 2012 was nearly 30 times smaller than that in 2008. This rapid decline in brightness is attributed to an unusually fast removal of dust from the system.

However, both models have unresolved issues.

One model relies on the presence of gas in the disk to cause aerodynamic drag, resulting in rapid accretion of material onto the star in a process termed runaway accretion. This mechanism requires a gas-to-dust ratio of at least unity, however, and it is not clear whether such a large quantity of gas could be either retained by the primordial disk in which the star formed, or replenished. More plausible is the collisional cascade model, in which gravitationally bound dust grains experience successive cratering or wholly destructive collisions that eventually yield grains small enough to be blown out of the system. This model predicts that a brightness drop of the level observed should occur on a timescale several times longer than the 2.5 years documented. However, refining the size distribution of the pre-event dust population could help to match this theory to the data. One benefit of this model is that it predicts cyclical behaviour, such that dust replenishment could be expected to yield an observable signature within two decades of the time of depletion.

Regardless of the uncertainties surrounding the models, the extremely rapid changes in this dusty system are certain to provoke further discussion of planetary-system evolution. Specific physical processes can usually be pinpointed from high-resolution images of disks that are close enough to Earth to have their structure — including gaps, rings and clumps¹ — spatially resolved. Such processes can also be uncovered through the disks' spectra¹¹, which shed light on the characteristics of the dust therein. Ideally, all of these data would be available for a source but, in this case, obtaining such information would have required advance knowledge of the event.

The disappearance of the excess infrared radiation from TYC 8241 2652 1 in less than two years is incredibly fast by our current understanding, and the impact of this is difficult to predict. The dust-clearing models proposed by Melis *et al.* could be refined to bring them more into line with conventional theory. And theories that have been developed for other stars and that were adapted to TYC 8241 2652 1 could be redeveloped. However, perhaps the most exciting possibility is

that the brightness drop represents a stage of terrestrial-planet formation that occurs so quickly that we have not been lucky enough to glimpse it until now. ■

Margaret Moerchen is at the *Leiden Observatory, 2333 CA Leiden, the Netherlands, and at the European Southern Observatory, Santiago, Chile.*
e-mail: mmoerche@eso.org

- Wyatt, M. C. *Annu. Rev. Astron. Astrophys.* **46**, 339–383 (2008).
- Melis, C. *et al. Nature* **487**, 74–76 (2012).
- Aumann, H. H. *et al. Astrophys. J.* **278**, L23–L27 (1984).
- Kenyon, S. J. & Bromley, B. C. *Astrophys. J. Suppl. Ser.* **179**, 451–483 (2008).
- Wyatt, M. C. *et al. Astrophys. J.* **658**, 569–583 (2007).
- Høg, E. *et al. Astron. Astrophys.* **355**, L27–L30 (2000).
- Neugebauer, G. *et al. Astrophys. J.* **278**, L1–L6 (1984).
- Ishihara, D. *et al. Astron. Astrophys.* **514**, A1 (2010).
- Wright, E. L. *et al. Astron. J.* **140**, 1868–1881 (2010).
- Telesco, C. M. *et al. Proc. SPIE* **3354**, 534–544 (1998).
- Lisse, C. M. *et al. Astrophys. J.* **701**, 2019–2032 (2009).

influence — commonness is rare. Consistent with this paradigm are neutral models of assemblage structure⁵, in which individuals of different species are seen as being effectively ecologically equivalent. These models also consider the process that determines which species become common and dominate assemblages to be essentially stochastic, or to have a strong stochastic component.

Only in their strictest form are these paradigms necessarily mutually exclusive. More usefully, they can be seen as providing complementary viewpoints that serve to shine a light on different components of assemblages. Indeed, many researchers find it useful to embed their work in one or other paradigm. Mi and colleagues effectively ask which paradigm is the more relevant.

Determining the ecological similarity of rare and common species is not straightforward, particularly when large numbers of species are involved. In principle, this similarity can be measured on numerous axes, including multiple components of morphology, physiology, life history and behaviour, but it is often not obvious a priori which axes might be the most important. Mi *et al.* take a short cut. They make the reasonable assumption that more closely related species (those that are genetically more similar) are on average more likely to share important traits, so that one can use the level of relatedness as a proxy for ecological similarity.

The authors used data from study plots of forest dynamics for their analysis. These plots, in which all individual trees (above a given threshold size) are painstakingly mapped, measured and identified, have made an extraordinary contribution to the study of

ECOLOGY

The importance of being rare

Uncovering the contribution of rare species to ecosystems is crucial to predicting the impacts of biodiversity loss. It seems that these species can be ecologically very different from their common relatives, but only in some cases.

KEVIN J. GASTON

Rare species are not invariably threatened with imminent extinction. However, those species that are threatened are almost invariably rare. The recent and projected future high rates of local, regional and global species loss have focused renewed attention on the need to understand the role of rare species in the structure and function of ecological assemblages. Writing in *American Naturalist*, Mi *et al.*¹ test the simple and fundamental question that lies at the heart of this understanding: namely, whether the rare species within a given site are ecologically dissimilar from the common ones.

Arguably, one can distinguish two prevailing paradigms in community ecology. What might be termed the 'rarity-richness' paradigm focuses on the notion that a crucial difference in the structure and function of broadly similar ecological assemblages (such as the plants, birds or mammals that inhabit different grasslands or tropical forests) is the number of species that they comprise, and, given that many

of these species will have few individuals, the number of rare species that are present^{2,3}. In this paradigm, rare species are seen as often having key roles in the structure and function of the assemblage, such as providing ecosystem goods and services, because they are ecologically different from common species. As a result, models of assemblage structure that are based on the concept of niche differentiation — an ecological process in which different species partition available resources between one another in space and/or time — are closely associated with this paradigm.

The second might be termed the 'commonness-dominance' paradigm. Here, attention is focused on the common species, rather than the rare ones, on the grounds that common species account for the vast majority of individuals, biomass and energy use in an assemblage, and therefore many of its essential characteristics⁴. In this paradigm, species richness, and the rare species that make up much of it, are considered less important than the composition of the common species, which, although relatively few in number, have disproportionate

However, both models have unresolved issues.

One model relies on the presence of gas in the disk to cause aerodynamic drag, resulting in rapid accretion of material onto the star in a process termed runaway accretion. This mechanism requires a gas-to-dust ratio of at least unity, however, and it is not clear whether such a large quantity of gas could be either retained by the primordial disk in which the star formed, or replenished. More plausible is the collisional cascade model, in which gravitationally bound dust grains experience successive cratering or wholly destructive collisions that eventually yield grains small enough to be blown out of the system. This model predicts that a brightness drop of the level observed should occur on a timescale several times longer than the 2.5 years documented. However, refining the size distribution of the pre-event dust population could help to match this theory to the data. One benefit of this model is that it predicts cyclical behaviour, such that dust replenishment could be expected to yield an observable signature within two decades of the time of depletion.

Regardless of the uncertainties surrounding the models, the extremely rapid changes in this dusty system are certain to provoke further discussion of planetary-system evolution. Specific physical processes can usually be pinpointed from high-resolution images of disks that are close enough to Earth to have their structure — including gaps, rings and clumps¹ — spatially resolved. Such processes can also be uncovered through the disks' spectra¹¹, which shed light on the characteristics of the dust therein. Ideally, all of these data would be available for a source but, in this case, obtaining such information would have required advance knowledge of the event.

The disappearance of the excess infrared radiation from TYC 8241 2652 1 in less than two years is incredibly fast by our current understanding, and the impact of this is difficult to predict. The dust-clearing models proposed by Melis *et al.* could be refined to bring them more into line with conventional theory. And theories that have been developed for other stars and that were adapted to TYC 8241 2652 1 could be redeveloped. However, perhaps the most exciting possibility is

that the brightness drop represents a stage of terrestrial-planet formation that occurs so quickly that we have not been lucky enough to glimpse it until now. ■

Margaret Moerchen is at the *Leiden Observatory, 2333 CA Leiden, the Netherlands, and at the European Southern Observatory, Santiago, Chile.*
e-mail: mmoerche@eso.org

- Wyatt, M. C. *Annu. Rev. Astron. Astrophys.* **46**, 339–383 (2008).
- Melis, C. *et al. Nature* **487**, 74–76 (2012).
- Aumann, H. H. *et al. Astrophys. J.* **278**, L23–L27 (1984).
- Kenyon, S. J. & Bromley, B. C. *Astrophys. J. Suppl. Ser.* **179**, 451–483 (2008).
- Wyatt, M. C. *et al. Astrophys. J.* **658**, 569–583 (2007).
- Høg, E. *et al. Astron. Astrophys.* **355**, L27–L30 (2000).
- Neugebauer, G. *et al. Astrophys. J.* **278**, L1–L6 (1984).
- Ishihara, D. *et al. Astron. Astrophys.* **514**, A1 (2010).
- Wright, E. L. *et al. Astron. J.* **140**, 1868–1881 (2010).
- Telesco, C. M. *et al. Proc. SPIE* **3354**, 534–544 (1998).
- Lisse, C. M. *et al. Astrophys. J.* **701**, 2019–2032 (2009).

influence — commonness is rare. Consistent with this paradigm are neutral models of assemblage structure⁵, in which individuals of different species are seen as being effectively ecologically equivalent. These models also consider the process that determines which species become common and dominate assemblages to be essentially stochastic, or to have a strong stochastic component.

Only in their strictest form are these paradigms necessarily mutually exclusive. More usefully, they can be seen as providing complementary viewpoints that serve to shine a light on different components of assemblages. Indeed, many researchers find it useful to embed their work in one or other paradigm. Mi and colleagues effectively ask which paradigm is the more relevant.

Determining the ecological similarity of rare and common species is not straightforward, particularly when large numbers of species are involved. In principle, this similarity can be measured on numerous axes, including multiple components of morphology, physiology, life history and behaviour, but it is often not obvious a priori which axes might be the most important. Mi *et al.* take a short cut. They make the reasonable assumption that more closely related species (those that are genetically more similar) are on average more likely to share important traits, so that one can use the level of relatedness as a proxy for ecological similarity.

The authors used data from study plots of forest dynamics for their analysis. These plots, in which all individual trees (above a given threshold size) are painstakingly mapped, measured and identified, have made an extraordinary contribution to the study of

ECOLOGY

The importance of being rare

Uncovering the contribution of rare species to ecosystems is crucial to predicting the impacts of biodiversity loss. It seems that these species can be ecologically very different from their common relatives, but only in some cases.

KEVIN J. GASTON

Rare species are not invariably threatened with imminent extinction. However, those species that are threatened are almost invariably rare. The recent and projected future high rates of local, regional and global species loss have focused renewed attention on the need to understand the role of rare species in the structure and function of ecological assemblages. Writing in *American Naturalist*, Mi *et al.*¹ test the simple and fundamental question that lies at the heart of this understanding: namely, whether the rare species within a given site are ecologically dissimilar from the common ones.

Arguably, one can distinguish two prevailing paradigms in community ecology. What might be termed the 'rarity-richness' paradigm focuses on the notion that a crucial difference in the structure and function of broadly similar ecological assemblages (such as the plants, birds or mammals that inhabit different grasslands or tropical forests) is the number of species that they comprise, and, given that many

of these species will have few individuals, the number of rare species that are present^{2,3}. In this paradigm, rare species are seen as often having key roles in the structure and function of the assemblage, such as providing ecosystem goods and services, because they are ecologically different from common species. As a result, models of assemblage structure that are based on the concept of niche differentiation — an ecological process in which different species partition available resources between one another in space and/or time — are closely associated with this paradigm.

The second might be termed the 'commonness-dominance' paradigm. Here, attention is focused on the common species, rather than the rare ones, on the grounds that common species account for the vast majority of individuals, biomass and energy use in an assemblage, and therefore many of its essential characteristics⁴. In this paradigm, species richness, and the rare species that make up much of it, are considered less important than the composition of the common species, which, although relatively few in number, have disproportionate

community ecology. Mi and colleagues drew data from a temperate and tropical network of 15 such plots, each of 16–52 hectares. Nine of the plots are gap-dominated forests, in which the fall of large trees removes shade, allowing other plants to compete to exploit the canopy gap. The other six are disturbance-dominated forests, shaped by factors such as typhoons and fires. The researchers used the density of individual species as the measurement of rarity, rather than other dimensions such as geographical range, and found evidence for niche differentiation — supporting the rarity–richness paradigm — between rare and common species, and among rare species themselves, in six of the nine gap-dominated forests. However, their analysis found no evidence for niche differentiation in the six disturbance-dominated and the three other gap-dominated forests.

One might argue that these results are simply inconclusive. However, there is no reason to assume that all assemblages will be structured in generally the same way, and that therefore the same ecological processes should apply. Indeed, the competition for access to canopy gaps that characterizes gap-dominated forests suggests that niche-differentiation mechanisms might be disproportionately important in these habitats. Similarly, in disturbance-dominated forests, stochastic processes might predominate, such that rare and common species are likely to have an increased probability of sharing traits because they are all faced with the vital need to deal with the effects of the disturbance.

A crucial issue that Mi and colleagues' analysis does not address is the influence of spatial scale on the observed ecological differentiation between rare and common species. Although some species are rare everywhere that they occur, locally common species are almost invariably rare in many of the places they occupy. Given that between-species variation in traits is typically much greater than within-species variation, this means that the same trait combinations will often be possessed by a species that is common in one place and rare in another. Thus, individual species are likely to contribute differently to the structure and function of ecological assemblages both across locations and at different scales.

A scenario in which some assemblages are better characterized by a rarity–richness paradigm and others by a commonness–dominance one, and thus in which rare species have different roles in different circumstances, presents a management challenge. Mi and colleagues' findings suggest that human interventions aimed at preserving ecosystem structure and function may need to vary the focus of attention on rare and common species in different systems. And it is likely that such approaches will need to be yet more flexible when addressing situations in which, for example, species' respective roles change through time. ■

Kevin J. Gaston is at the Environment and Sustainability Institute, University of Exeter, Penryn TR10 9EZ, UK.
e-mail: k.j.gaston@exeter.ac.uk

1. Mi, X. *et al.* *Am. Nat.* **180**, E17–E30 (2012).

2. Duffy, J. E. *Front. Ecol. Environ.* **7**, 437–444 (2009).
3. Cardinale, B. J. *et al.* *Nature* **486**, 59–67 (2012).
4. Gaston, K. J. *BioScience* **61**, 354–362 (2011).
5. Hubbell, S. P. *The Unified Neutral Theory of Biodiversity and Biogeography* (Princeton Univ. Press, 2001).

MICROBIOLOGY

Fat, bile and gut microbes

Western-style diets could be contributing to the rapid increase in inflammatory bowel disease. New research suggests that dietary fat can alter bile composition and so favour the growth of pro-inflammatory gut microbes. [SEE LETTER P.104](#)

PETER J. TURNBAUGH

Humans never truly dine alone — our diet is intimately linked to the functioning of the trillions of microbes that inhabit our gastrointestinal tract, the gut microbiota. These microorganisms help us to digest complex carbohydrates and other substances that cannot be directly metabolized by our cells' own enzymes¹, and our resident microbes are, in turn, affected by what we eat^{2,3}. However, it is unclear how exactly our dietary choices influence our microbiota and what implications this could have for disease. On page 104 of this issue, Devkota *et al.*⁴ present a pioneering study that demonstrates that dietary fat can alter the gut microbiota of mice indirectly by changing the animals' pool of bile acids — steroids that are produced by the liver and secreted into the intestine. The authors also provide a plausible mechanism through which diets that are high in certain saturated fats might contribute to inflammatory bowel disease.

Devkota *et al.* set out to explore whether and, if so, how, the animal-derived saturated

fats that are abundant in Western diets could affect gut microbes in mice. They tested three sterilized high-fat diets, two of which contained saturated fats from either milk or lard, and a third that was based on safflower oil (a polyunsaturated fat). The researchers then analysed the diversity of bacteria in the mouse gut by sequencing the microbial gene that encodes 16S ribosomal RNA, a highly conserved molecule with a sequence that is used by researchers to classify bacteria.

The milk-fat diet promoted an increase in the abundance of a microbe known as *Bilophila wadsworthia* and, to a lesser extent, of others such as *Lactobacillus murinus* and species of *Alistipes*. Notably, *B. wadsworthia* is found in small numbers in the guts of healthy people, but is often detected in clinical samples from patients with appendicitis and other infections⁵. Although the consumption of milk fat did not affect the health of wild-type mice, it did increase the onset and incidence of colitis in genetically susceptible mice that lacked interleukin-10, an anti-inflammatory signalling molecule.

The authors then studied interleukin-10-

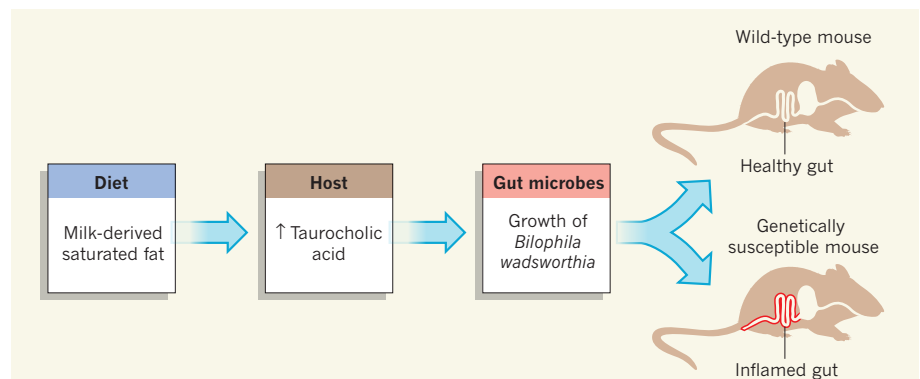


Figure 1 | Dietary road to colitis. Devkota *et al.*⁴ report that, in mice, a diet rich in milk-derived saturated fat led to increased liver production of taurocholic acid (a substance that aids the digestion of fats), which, in turn, stimulated the growth of certain gut bacteria. In genetically susceptible mice, one of these microbial species (*Bilophila wadsworthia*) triggered intestinal inflammation and colitis.

community ecology. Mi and colleagues drew data from a temperate and tropical network of 15 such plots, each of 16–52 hectares. Nine of the plots are gap-dominated forests, in which the fall of large trees removes shade, allowing other plants to compete to exploit the canopy gap. The other six are disturbance-dominated forests, shaped by factors such as typhoons and fires. The researchers used the density of individual species as the measurement of rarity, rather than other dimensions such as geographical range, and found evidence for niche differentiation — supporting the rarity–richness paradigm — between rare and common species, and among rare species themselves, in six of the nine gap-dominated forests. However, their analysis found no evidence for niche differentiation in the six disturbance-dominated and the three other gap-dominated forests.

One might argue that these results are simply inconclusive. However, there is no reason to assume that all assemblages will be structured in generally the same way, and that therefore the same ecological processes should apply. Indeed, the competition for access to canopy gaps that characterizes gap-dominated forests suggests that niche-differentiation mechanisms might be disproportionately important in these habitats. Similarly, in disturbance-dominated forests, stochastic processes might predominate, such that rare and common species are likely to have an increased probability of sharing traits because they are all faced with the vital need to deal with the effects of the disturbance.

A crucial issue that Mi and colleagues' analysis does not address is the influence of spatial scale on the observed ecological differentiation between rare and common species. Although some species are rare everywhere that they occur, locally common species are almost invariably rare in many of the places they occupy. Given that between-species variation in traits is typically much greater than within-species variation, this means that the same trait combinations will often be possessed by a species that is common in one place and rare in another. Thus, individual species are likely to contribute differently to the structure and function of ecological assemblages both across locations and at different scales.

A scenario in which some assemblages are better characterized by a rarity–richness paradigm and others by a commonness–dominance one, and thus in which rare species have different roles in different circumstances, presents a management challenge. Mi and colleagues' findings suggest that human interventions aimed at preserving ecosystem structure and function may need to vary the focus of attention on rare and common species in different systems. And it is likely that such approaches will need to be yet more flexible when addressing situations in which, for example, species' respective roles change through time. ■

Kevin J. Gaston is at the Environment and Sustainability Institute, University of Exeter, Penryn TR10 9EZ, UK.
e-mail: k.j.gaston@exeter.ac.uk

1. Mi, X. *et al.* *Am. Nat.* **180**, E17–E30 (2012).

2. Duffy, J. E. *Front. Ecol. Environ.* **7**, 437–444 (2009).
3. Cardinale, B. J. *et al.* *Nature* **486**, 59–67 (2012).
4. Gaston, K. J. *BioScience* **61**, 354–362 (2011).
5. Hubbell, S. P. *The Unified Neutral Theory of Biodiversity and Biogeography* (Princeton Univ. Press, 2001).

MICROBIOLOGY

Fat, bile and gut microbes

Western-style diets could be contributing to the rapid increase in inflammatory bowel disease. New research suggests that dietary fat can alter bile composition and so favour the growth of pro-inflammatory gut microbes. [SEE LETTER P.104](#)

PETER J. TURNBAUGH

Humans never truly dine alone — our diet is intimately linked to the functioning of the trillions of microbes that inhabit our gastrointestinal tract, the gut microbiota. These microorganisms help us to digest complex carbohydrates and other substances that cannot be directly metabolized by our cells' own enzymes¹, and our resident microbes are, in turn, affected by what we eat^{2,3}. However, it is unclear how exactly our dietary choices influence our microbiota and what implications this could have for disease. On page 104 of this issue, Devkota *et al.*⁴ present a pioneering study that demonstrates that dietary fat can alter the gut microbiota of mice indirectly by changing the animals' pool of bile acids — steroids that are produced by the liver and secreted into the intestine. The authors also provide a plausible mechanism through which diets that are high in certain saturated fats might contribute to inflammatory bowel disease.

Devkota *et al.* set out to explore whether and, if so, how, the animal-derived saturated

fats that are abundant in Western diets could affect gut microbes in mice. They tested three sterilized high-fat diets, two of which contained saturated fats from either milk or lard, and a third that was based on safflower oil (a polyunsaturated fat). The researchers then analysed the diversity of bacteria in the mouse gut by sequencing the microbial gene that encodes 16S ribosomal RNA, a highly conserved molecule with a sequence that is used by researchers to classify bacteria.

The milk-fat diet promoted an increase in the abundance of a microbe known as *Bilophila wadsworthia* and, to a lesser extent, of others such as *Lactobacillus murinus* and species of *Alistipes*. Notably, *B. wadsworthia* is found in small numbers in the guts of healthy people, but is often detected in clinical samples from patients with appendicitis and other infections⁵. Although the consumption of milk fat did not affect the health of wild-type mice, it did increase the onset and incidence of colitis in genetically susceptible mice that lacked interleukin-10, an anti-inflammatory signalling molecule.

The authors then studied interleukin-10-

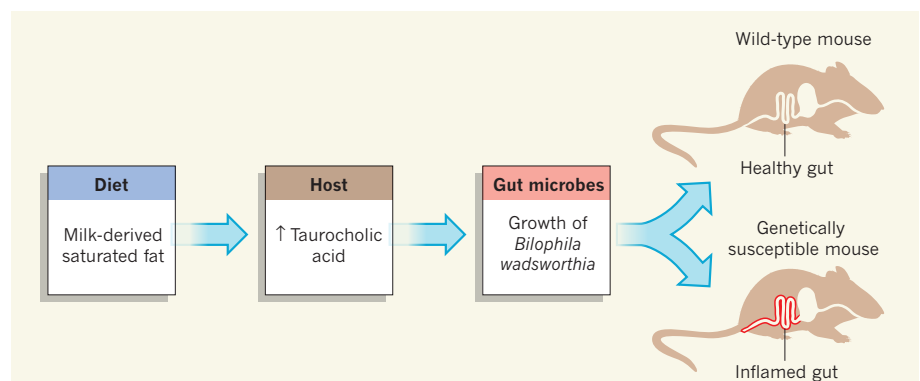


Figure 1 | Dietary road to colitis. Devkota *et al.*⁴ report that, in mice, a diet rich in milk-derived saturated fat led to increased liver production of taurocholic acid (a substance that aids the digestion of fats), which, in turn, stimulated the growth of certain gut bacteria. In genetically susceptible mice, one of these microbial species (*Bilophila wadsworthia*) triggered intestinal inflammation and colitis.

deficient mice that had been raised in germ-free conditions and thus harboured no microbes in their guts. After feeding these mice either a low- or a high-fat diet, Devkota *et al.* gave the animals solutions containing different microbes (*B. wadsworthia*, *L. murinus* or *Alistipes*). Only the mice that had consumed milk fat were successfully colonized by *B. wadsworthia*, and only these animals developed colitis.

Why did milk fat stimulate the growth of *B. wadsworthia*? This microbe thrives in the presence of taurocholic acid (a combination of a bile acid and taurine, a sulphur-containing molecule)⁶, and taurocholic acid is thought to be produced more readily during the ingestion of milk fats⁷. Indeed, the researchers confirmed that milk-fat consumption led to an increased concentration of taurocholic acid in the bile of interleukin-10-deficient mice, and that this bile promoted the growth of *B. wadsworthia* *in vitro*. Furthermore, the addition of taurocholic acid to a low-fat diet stimulated *B. wadsworthia* colonization of interleukin-10-deficient mice — irrespective of whether the animals were germ-free — and prompted the development of colitis.

Devkota and colleagues' results support a model in which a 'perfect storm' of milk-fat consumption, shifts in bile-acid production, aberrant growth of pro-inflammatory microbes and genetic susceptibility conspire to produce colitis (Fig. 1). Consistent with this hypothesis, previous studies in humans^{8,9} have demonstrated an association between inflammatory bowel disease and sulphate-reducing bacteria — which, like *B. wadsworthia*, can metabolize certain sulphur-containing compounds. The by-products of this bacterial metabolism, namely hydrogen sulphide (H₂S), might disrupt the epithelial tissue that lines the inside of the gut and that acts as a barrier against pathogens and toxins. Such disruption could magnify the effects of certain immune responses to antigens, of pro-inflammatory compounds produced by the gut microbiota, and of viral infections¹⁰.

The present study highlights the key role that bile acids can have in shaping the gut microbiota. However, more work is needed to find out whether and how the bile-acid pool is affected by various perturbations such as dietary changes, gastric-bypass surgery, probiotics (microbes that can confer health benefits) and therapeutic drugs. It will also be important to determine how gut microbes can resist or even metabolize bile acids, and to what extent inter-individual variation in microbial community structure and function influences the effects of potentially damaging microbes such as *B. wadsworthia* on predisposition to disease.

More broadly, Devkota and colleagues' work emphasizes the importance of viewing nutrition from a perspective that encompasses both our human and microbial genomes. As the authors elegantly show, diets that provide the same number of calories — in total and from fat — can have remarkably different

effects depending on the type of fat. It will be interesting to identify the specific components of our diet that influence microbial community structure, and to find out whether diets rich in saturated fat can drive the expansion of *B. wadsworthia* or other potentially harmful microbes in humans.

Importantly, the effects of a given diet on the host and on the gut microbiota are not always independent; in this study, diet-driven changes in the production of bile acids affected gut microbes that, in turn, triggered disease. This line of research could ultimately lead to dietary recommendations tailored to match the idiosyncrasies of each person's gut microbiota. Although we may be years away from achieving this goal, we should at least consider the many intended and unintended consequences that our diet has on our microbial dinner guests. ■

ORGANIC CHEMISTRY

Reactions at the end of a tether

A remarkable reaction that reverses the chemical behaviour of molecules known as 1,3-diketones allows a new strategy that could be used to prepare a range of potentially useful, naturally occurring compounds. SEE LETTER P.86

STEFAN ROESNER & VARINDER K. AGGARWAL

Building architecturally complex organic molecules from simple building blocks is akin to building the Taj Mahal from rough chunks of marble — with breathtaking results arising from unassuming starting materials. The process often involves making molecular building blocks called nucleophiles and electrophiles, which attract each other. Once pulled together by attraction, the blocks can react with each other, creating a new, more complex molecule. On page 86 of this issue, Leighton and colleagues¹ describe some clever molecular choreography that enables them to convert a nucleophile into an electrophile, so that it can react with another nucleophilic partner. Not only is this work of fundamental scientific interest, but it also paves the way for short synthetic routes of structurally complex molecules*.

Some of the most commonly used nucleophiles in organic synthesis are organometallic reagents — compounds that contain a carbon–metal bond. Organometallic reagents that contain allyl groups (CH₂CH=CH₂) are particularly useful. These react with aldehydes to form 'homoallylic' alcohols, which contain carbon–carbon double bonds — useful starting points

*This article and the paper¹ under discussion were published online on 27 June 2012.

Peter J. Turnbaugh is at the FAS Center for Systems Biology, Harvard University, Cambridge, Massachusetts 02138, USA. e-mail: pturnbaugh@fas.harvard.edu

1. Turnbaugh, P. J., Henrissat, B. & Gordon, J. I. *Acta Crystallogr. F* **66**, 1261–1264 (2010).
2. Turnbaugh, P. J. *et al. Sci. Transl. Med.* **1**, 6ra14 (2009).
3. Walker, A. W. *et al. ISME J.* **5**, 220–230 (2011).
4. Devkota, S. *et al. Nature* **487**, 104–108 (2012).
5. Baron, E. J. *Anaerobe* **3**, 83–86 (1997).
6. Laue, H., Denger, K. & Cook, A. M. *Appl. Environ. Microbiol.* **63**, 2016–2021 (1997).
7. Lindstedt, S., Avigan, J., Goodman, D. S., Sjövall, J. & Steinberg, D. J. *Clin. Invest.* **44**, 1754–1765 (1965).
8. Loubinoux, J., Bronowicki, J. P., Pereira, I. A., Mouguel, J. L. & Faou, A. E. *FEMS Microbiol. Ecol.* **40**, 107–112 (2002).
9. Rowan, F. E., Docherty, N. G., Coffey, J. C. & O'Connell, P. R. *Br. J. Surg.* **96**, 151–158 (2009).
10. Cadwell, K. *et al. Cell* **141**, 1135–1145 (2010).

for a wealth of other synthetic reactions. Perhaps the most widely used allylic nucleophiles are allylboranes, in which the allyl group is attached to a boron atom, and allylsilanes, in which the allyl group is attached to silicon (Fig. 1a).

In many cases, the alcohols that form are chiral — that is, they can exist as two mirror-image isomers known as enantiomers. Chiral variants of boranes have long been used to control which enantiomer of a product forms, making these reactions a cornerstone of organic synthesis. The US chemist Herbert C. Brown won a Nobel prize for his work developing chiral boranes^{2–4}, many of which are still used for the synthesis of complex molecules more than 30 years after their discovery. More recently, Leighton's group has reported⁵ that the members of a family of chiral allylsilanes (Fig. 1a) react with aldehydes to yield products with even higher enantioselectivity than can be achieved using Brown's reagents. These silanes are increasingly being used for synthesis.

Some of the most interesting targets of organic synthesis are compounds that contain carbonyl groups (C=O) separated from homoallylic alcohols by a single carbon atom. Such motifs could, in principle, be assembled using the strategies discussed above, except that the starting material would not be a simple aldehyde, but a 1,3-diketone — a

deficient mice that had been raised in germ-free conditions and thus harboured no microbes in their guts. After feeding these mice either a low- or a high-fat diet, Devkota *et al.* gave the animals solutions containing different microbes (*B. wadsworthia*, *L. murinus* or *Alistipes*). Only the mice that had consumed milk fat were successfully colonized by *B. wadsworthia*, and only these animals developed colitis.

Why did milk fat stimulate the growth of *B. wadsworthia*? This microbe thrives in the presence of taurocholic acid (a combination of a bile acid and taurine, a sulphur-containing molecule)⁶, and taurocholic acid is thought to be produced more readily during the ingestion of milk fats⁷. Indeed, the researchers confirmed that milk-fat consumption led to an increased concentration of taurocholic acid in the bile of interleukin-10-deficient mice, and that this bile promoted the growth of *B. wadsworthia* *in vitro*. Furthermore, the addition of taurocholic acid to a low-fat diet stimulated *B. wadsworthia* colonization of interleukin-10-deficient mice — irrespective of whether the animals were germ-free — and prompted the development of colitis.

Devkota and colleagues' results support a model in which a 'perfect storm' of milk-fat consumption, shifts in bile-acid production, aberrant growth of pro-inflammatory microbes and genetic susceptibility conspire to produce colitis (Fig. 1). Consistent with this hypothesis, previous studies in humans^{8,9} have demonstrated an association between inflammatory bowel disease and sulphate-reducing bacteria — which, like *B. wadsworthia*, can metabolize certain sulphur-containing compounds. The by-products of this bacterial metabolism, namely hydrogen sulphide (H₂S), might disrupt the epithelial tissue that lines the inside of the gut and that acts as a barrier against pathogens and toxins. Such disruption could magnify the effects of certain immune responses to antigens, of pro-inflammatory compounds produced by the gut microbiota, and of viral infections¹⁰.

The present study highlights the key role that bile acids can have in shaping the gut microbiota. However, more work is needed to find out whether and how the bile-acid pool is affected by various perturbations such as dietary changes, gastric-bypass surgery, probiotics (microbes that can confer health benefits) and therapeutic drugs. It will also be important to determine how gut microbes can resist or even metabolize bile acids, and to what extent inter-individual variation in microbial community structure and function influences the effects of potentially damaging microbes such as *B. wadsworthia* on predisposition to disease.

More broadly, Devkota and colleagues' work emphasizes the importance of viewing nutrition from a perspective that encompasses both our human and microbial genomes. As the authors elegantly show, diets that provide the same number of calories — in total and from fat — can have remarkably different

effects depending on the type of fat. It will be interesting to identify the specific components of our diet that influence microbial community structure, and to find out whether diets rich in saturated fat can drive the expansion of *B. wadsworthia* or other potentially harmful microbes in humans.

Importantly, the effects of a given diet on the host and on the gut microbiota are not always independent; in this study, diet-driven changes in the production of bile acids affected gut microbes that, in turn, triggered disease. This line of research could ultimately lead to dietary recommendations tailored to match the idiosyncrasies of each person's gut microbiota. Although we may be years away from achieving this goal, we should at least consider the many intended and unintended consequences that our diet has on our microbial dinner guests. ■

ORGANIC CHEMISTRY

Reactions at the end of a tether

A remarkable reaction that reverses the chemical behaviour of molecules known as 1,3-diketones allows a new strategy that could be used to prepare a range of potentially useful, naturally occurring compounds. SEE LETTER P.86

STEFAN ROESNER & VARINDER K. AGGARWAL

Building architecturally complex organic molecules from simple building blocks is akin to building the Taj Mahal from rough chunks of marble — with breathtaking results arising from unassuming starting materials. The process often involves making molecular building blocks called nucleophiles and electrophiles, which attract each other. Once pulled together by attraction, the blocks can react with each other, creating a new, more complex molecule. On page 86 of this issue, Leighton and colleagues¹ describe some clever molecular choreography that enables them to convert a nucleophile into an electrophile, so that it can react with another nucleophilic partner. Not only is this work of fundamental scientific interest, but it also paves the way for short synthetic routes of structurally complex molecules*.

Some of the most commonly used nucleophiles in organic synthesis are organometallic reagents — compounds that contain a carbon–metal bond. Organometallic reagents that contain allyl groups (CH₂CH=CH₂) are particularly useful. These react with aldehydes to form 'homoallylic' alcohols, which contain carbon–carbon double bonds — useful starting points

*This article and the paper¹ under discussion were published online on 27 June 2012.

Peter J. Turnbaugh is at the FAS Center for Systems Biology, Harvard University, Cambridge, Massachusetts 02138, USA. e-mail: pturnbaugh@fas.harvard.edu

1. Turnbaugh, P. J., Henrissat, B. & Gordon, J. I. *Acta Crystallogr. F* **66**, 1261–1264 (2010).
2. Turnbaugh, P. J. *et al. Sci. Transl. Med.* **1**, 6ra14 (2009).
3. Walker, A. W. *et al. ISME J.* **5**, 220–230 (2011).
4. Devkota, S. *et al. Nature* **487**, 104–108 (2012).
5. Baron, E. J. *Anaerobe* **3**, 83–86 (1997).
6. Laue, H., Denger, K. & Cook, A. M. *Appl. Environ. Microbiol.* **63**, 2016–2021 (1997).
7. Lindstedt, S., Avigan, J., Goodman, D. S., Sjövall, J. & Steinberg, D. J. *Clin. Invest.* **44**, 1754–1765 (1965).
8. Loubinoux, J., Bronowicki, J. P., Pereira, I. A., Mouguel, J. L. & Faou, A. E. *FEMS Microbiol. Ecol.* **40**, 107–112 (2002).
9. Rowan, F. E., Docherty, N. G., Coffey, J. C. & O'Connell, P. R. *Br. J. Surg.* **96**, 151–158 (2009).
10. Cadwell, K. *et al. Cell* **141**, 1135–1145 (2010).

for a wealth of other synthetic reactions. Perhaps the most widely used allylic nucleophiles are allylboranes, in which the allyl group is attached to a boron atom, and allylsilanes, in which the allyl group is attached to silicon (Fig. 1a).

In many cases, the alcohols that form are chiral — that is, they can exist as two mirror-image isomers known as enantiomers. Chiral variants of boranes have long been used to control which enantiomer of a product forms, making these reactions a cornerstone of organic synthesis. The US chemist Herbert C. Brown won a Nobel prize for his work developing chiral boranes^{2–4}, many of which are still used for the synthesis of complex molecules more than 30 years after their discovery. More recently, Leighton's group has reported⁵ that the members of a family of chiral allylsilanes (Fig. 1a) react with aldehydes to yield products with even higher enantioselectivity than can be achieved using Brown's reagents. These silanes are increasingly being used for synthesis.

Some of the most interesting targets of organic synthesis are compounds that contain carbonyl groups (C=O) separated from homoallylic alcohols by a single carbon atom. Such motifs could, in principle, be assembled using the strategies discussed above, except that the starting material would not be a simple aldehyde, but a 1,3-diketone — a

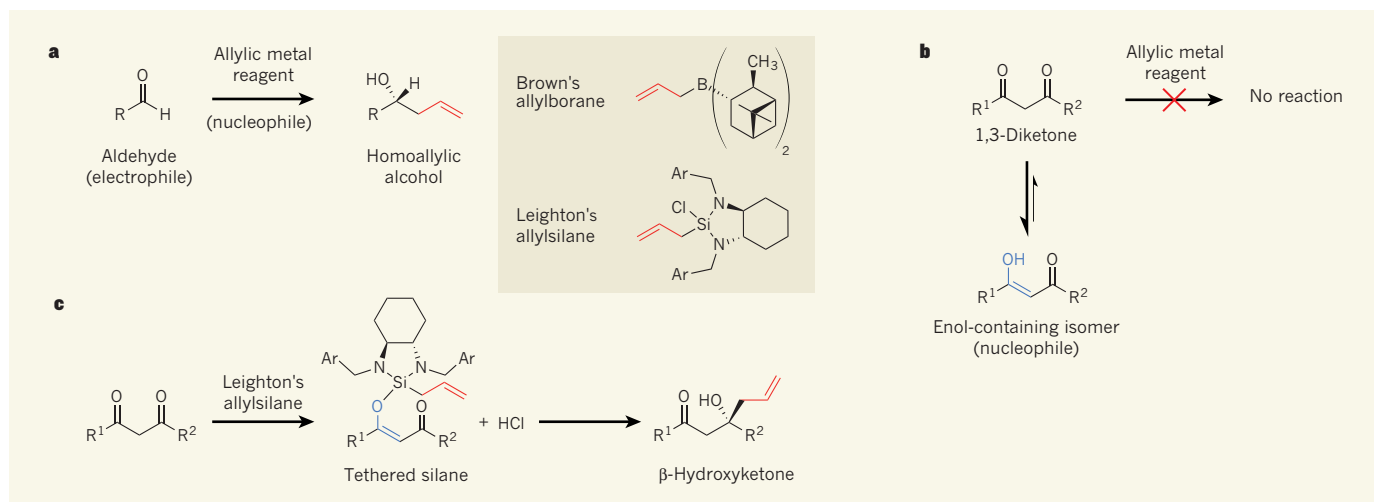


Figure 1 | Allylation reactions of aldehydes and 1,3-diketones. **a**, Electrophilic aldehydes react with nucleophilic, chiral allylic metal reagents, such as the allylborane and allylsilane shown, to form homoallylic alcohols as a single mirror-image isomer (enantiomer). The allyl group transferred in the reaction is shown in red. R is an alkyl group (a saturated hydrocarbon); Ar is an aryl (benzene-containing) group. **b**, Unlike aldehydes, 1,3-diketones are nucleophilic, and so do not react with allylic metal reagents. This is because

they exist predominantly as enol-containing isomers; the enol group is shown in blue. R groups are alkyl or aryl groups. **c**, Leighton and colleagues¹ have reacted 1,3-diketones with an allylsilane⁵ previously developed in their group. The silane becomes tethered to the enol group and then takes part in an intramolecular reaction that generates a β-hydroxyketone as a single enantiomer. The tethering reaction generates HCl, an acid, which helps to promote the intramolecular reaction.

compound in which two carbonyl groups are separated by a single carbon atom (Fig. 1b). This is problematic, because such compounds are nucleophilic rather than electrophilic. Indeed, 1,3-dicarbonyl compounds (such as 1,3-diketones) are such valuable nucleophiles that chemists often append an additional carbonyl group to lone carbonyls in synthetic intermediates to enhance the lone carbonyl's nucleophilicity for subsequent reactions.

This nucleophilicity of 1,3-diketones is a consequence of tautomerism, a phenomenon in which the compounds rapidly interconvert between an electrophilic dicarbonyl isomer and a nucleophilic isomer that contains one carbonyl group and an enol group (HO–C=C; Fig. 1b). In fact, the equilibrium state of the interconversion process ensures that most of the molecules are enol-containing isomers. So how does one coerce these isomers to behave as electrophiles? It should be possible given that they still contain a carbonyl group, albeit one that is much less electrophilic than a lone carbonyl group because of electron donation from the enol isomer.

Leighton and colleagues' solution was to tether an allylsilane to the enol group of a 1,3-diketone (Fig. 1c). The allyl group could then attack the remaining carbonyl group in an intramolecular reaction to form an alcohol — the only drawback being that, in general, such reactions are extremely slow. The authors expected this sluggishness to be offset to some degree by the fact that the reaction is intramolecular, because intramolecular reactions are generally faster than intermolecular ones. Nevertheless, they didn't anticipate that this effect alone would be enough to promote the reaction.

However, the tethering process generates acid molecules, which are known to activate the nucleophilic attack of allyl silanes on

carbonyl groups⁶. Would this help the reaction to occur? When the authors combined 1,3-diketones with one of Leighton's allylsilanes, hey presto: they obtained homoallylic alcohols bearing a neighbouring carbonyl in high yield, and with exquisite control of the enantiomer formed during the reaction.

Leighton and colleagues went on to show that their reaction works well for a range of symmetrical 1,3-diketones. More impressively, in the case of unsymmetrical 1,3-diketones that have an alkyl group (a saturated hydrocarbon) on one side of the two carbonyls, and an aryl group (one containing a benzene ring or a benzene-like ring) on the other, the allylsilane selectively attacks the carbonyl adjacent to the alkyl. The authors observed that the reaction also discriminates between the two carbonyl groups in 1,3-diketones bearing two different aryl groups, but not between those in diketones bearing two different alkyl groups. This limitation represents a crucial challenge remaining to be solved, because reactions of dialkyl diketones are likely to be of greatest use in the synthesis of complex molecules.

The synthetic utility of the process is further enhanced by the ready availability of 1,3-diketones and of Leighton's allylsilane reagents (which are commercially available), and because the protocol required for the reaction is simple. Ultimately, it would be useful if the reactions could be conducted with chiral catalysts rather than chiral reagents, because catalysts are used in small, sub-stoichiometric quantities and so are more suitable for use in large-scale reactions.

The ability to efficiently design, synthesize and manufacture ever more complex molecules that have particular shapes and functions is important to the future needs of

society. Leighton and colleagues' work is a considerable step towards this goal, and, as a master synthetic chemist⁷, Leighton is well placed to apply this methodology to the challenging targets presented by nature. Such application will provide the definitive challenge to test these impressive reactions. ■

Stefan Roesner and Varinder K. Aggarwal
are at the School of Chemistry, University of Bristol, Bristol BS8 1TS, UK.
e-mail: v.aggarwal@bristol.ac.uk

1. Chalfoux, W. A., Reznik, S. K. & Leighton, J. L. *Nature* **487**, 86–89 (2012).
2. Brown, H. C. & Jadhav, P. K. *J. Am. Chem. Soc.* **105**, 2092–2093 (1983).
3. Jadhav, P. K., Bhat, K. S., Perumal, P. T. & Brown, H. C. *J. Org. Chem.* **51**, 432–439 (1986).
4. Brown, H. C., Randad, R. S., Bhat, K. S., Zaidlewicz, M. & Racherla, U. S. *J. Am. Chem. Soc.* **112**, 2389–2392 (1990).
5. Kubota, K. & Leighton, J. L. *Angew. Chem. Int. Edn* **42**, 946–948 (2003).
6. Kim, H., Ho, S. & Leighton, J. L. *J. Am. Chem. Soc.* **133**, 6517–6520 (2011).
7. Harrison, T. J., Ho, S. & Leighton, J. L. *J. Am. Chem. Soc.* **133**, 7308–7311 (2011).

CORRECTION

In the News & Views article 'Applied chemistry: Molecules meet materials' by Thomas E. Mallouk (*Nature* **485**, 450–451; 2012), the words "anode" and "cathode" were incorrectly interchanged in one of the sentences. This should have read: "In these conductors, many of the electrons and holes that are created by light absorption annihilate each other before they can escape from the pore network to the cell's anode and cathode, respectively, to allow electric current to be generated."

Neural population dynamics during reaching

Mark M. Churchland^{1,2,3*}, John P. Cunningham^{4,5*}, Matthew T. Kaufman^{2,3}, Justin D. Foster², Paul Nuyujukian^{6,7}, Stephen I. Ryu^{2,8} & Krishna V. Shenoy^{2,3,6,9}

Most theories of motor cortex have assumed that neural activity represents movement parameters. This view derives from what is known about primary visual cortex, where neural activity represents patterns of light. Yet it is unclear how well the analogy between motor and visual cortex holds. Single-neuron responses in motor cortex are complex, and there is marked disagreement regarding which movement parameters are represented. A better analogy might be with other motor systems, where a common principle is rhythmic neural activity. Here we find that motor cortex responses during reaching contain a brief but strong oscillatory component, something quite unexpected for a non-periodic behaviour. Oscillation amplitude and phase followed naturally from the preparatory state, suggesting a mechanistic role for preparatory neural activity. These results demonstrate an unexpected yet surprisingly simple structure in the population response. This underlying structure explains many of the confusing features of individual neural responses.

Motor and premotor cortex were among the first cortical areas to be extensively studied¹. Yet their basic response properties are poorly understood, and it remains controversial whether neural activity relates to muscles or to abstract movement features^{2–13}. At the heart of this debate is the complexity of individual neural responses, which exhibit a great variety of multiphasic patterns^{4,14,15}. One explanation is that responses represent many movement parameters:

$$r_n(t) = f_n(\text{param}_1(t), \text{param}_2(t), \text{param}_3(t) \dots) \quad (1)$$

where $r_n(t)$ is the firing rate of neuron n at time t , f_n is a tuning function, and $\text{param}_1(t)$, $\text{param}_2(t) \dots$ are arguments such as hand velocity or target position. Alternatively, motor cortex may constitute a dynamical system that generates and controls movement^{4,8,14–17}. In its simplest, deterministic form this can be expressed as:

$$\dot{\mathbf{r}}(t) = f(\mathbf{r}(t)) + \mathbf{u}(t) \quad (2)$$

where \mathbf{r} is a vector describing the firing rate of all neurons (the ‘population response’ or ‘neural state’), $\dot{\mathbf{r}}$ is its derivative, f is an unknown function, and \mathbf{u} is an external input. In this conception, neural responses reflect underlying dynamics and display ‘tuning’ only incidentally^{18,19}. If so, then dynamical features should be present in the population response. In looking for dynamical structure, we focused on a common principle for movement generation across the animal kingdom: the production of rhythmic, oscillatory activity^{20–22}.

Rhythmic responses in different systems

We first examined neural responses in a context where rhythmic pattern generation is known to occur. The medicinal leech generates rhythmic muscle contractions at ~ 1.5 Hz during swimming²³, and many single neurons display firing rate oscillations at that frequency (Fig. 1a)^{24,25}. Rhythmic structure was also present for cortical responses in the walking monkey: ~ 1 Hz oscillations matching the ~ 1 Hz movement of the arm (Fig. 1b). If single-neuron oscillations

are generated by population-level dynamics, then the population response (the neural state) should rotate with time¹⁵, much as the state of a pendulum rotates in the space defined by velocity and position. We projected the population response onto a two-dimensional state space and found rotations of the neural state for both the swimming leech (Fig. 1d; projection of 164 simultaneously recorded neurons) and the walking monkey (Fig. 1e; projection of 32 simultaneously recorded channels; also see Supplementary Movie 1). These observations, although not trivial, are largely expected for a neural dynamical system that generates rhythmic output²².

The projections in Fig. 1d, e were obtained via two steps. The first was the application of principal component analysis (PCA) to the population response. Inconveniently, PCA does not find dimensions relevant to dynamical structure. We therefore used a novel method that finds an informative plane within the top principal components (PCs). To be conservative, this ‘jPCA’ method was applied only to the top six PCs, which contain the six response patterns most strongly present in the data. The mathematical underpinnings regarding jPCA are described below, but the following is critical. Application of jPCA results in six jPCs: an orthonormal basis that spans exactly the same space as the first six PCs (Supplementary Movie 2). The first two jPCs capture the strongest rotational tendency in the data. The jPC projections are simply linear projections of response patterns that are strongly present in the data; if a given pattern is not present in the top six PCs it cannot be present in the jPCs.

The central finding of this study is that quasi-oscillatory neural responses are present during reaches. This is illustrated by the average firing rate of an example motor cortex neuron (Fig. 1c) and the corresponding population-level projection (Fig. 1f). The rotation of the neural state is short lived (~ 1 cycle) but otherwise resembles rotations seen during rhythmic movement. This finding is surprising—the reaches themselves are not rhythmic—yet it agrees with recent theoretical suggestions^{15,22}. There might be a concern that the patterns in

¹Department of Neuroscience, Kavli Institute for Brain Science, David Mahoney Center, Columbia University Medical Center, New York, New York 10032, USA. ²Department of Electrical Engineering, Stanford University, Stanford, California 94305, USA. ³Neurosciences Program, Stanford University, Stanford, California 94305, USA. ⁴Department of Biomedical Engineering, Washington University in St. Louis, St. Louis, Missouri 63130, USA. ⁵Department of Engineering, University of Cambridge, Cambridge CB2 1PZ, UK. ⁶Department of Bioengineering, Stanford University, Stanford, California 94305, USA. ⁷Stanford University School of Medicine, Stanford, California 94305, USA. ⁸Department of Neurosurgery, Palo Alto Medical Foundation, Palo Alto, California 94301, USA. ⁹Department of Neurobiology, Stanford University School of Medicine, Stanford, California 94305, USA.

*These authors contributed equally to this work.

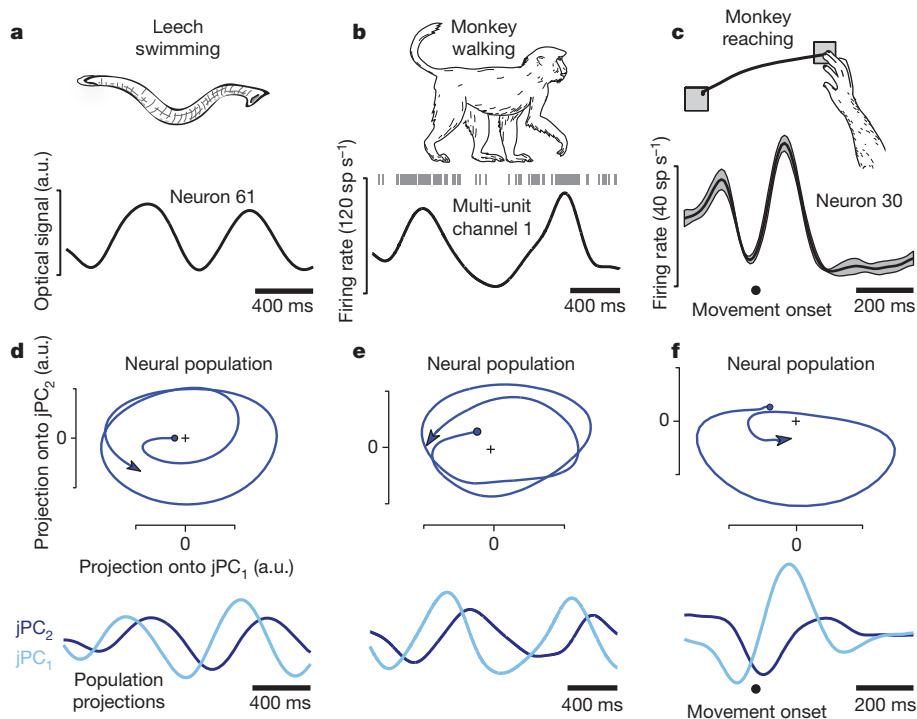


Figure 1 | Oscillation of neural firing rates during three movement types. **a**, Response of 1 of 164 neurons (simultaneously recorded using voltage-sensitive dye) in the isolated leech central nervous system during a swimming motor pattern. Responses (not averaged across repetitions) were filtered with a 100 ms Gaussian kernel. a.u., arbitrary units. **b**, Multi-unit response from 1 of 96 electrodes implanted in the arm representation of caudal premotor cortex. Data from 32 such channels were wirelessly transmitted during walking. sp s⁻¹, spikes per second. Responses (not averaged across repetitions) were filtered

with a 100 ms Gaussian kernel. **c**, Response of 1 of 118 neurons recorded from motor cortex of a reaching monkey (N) using single-electrode techniques. Firing rates were smoothed with a 24 ms Gaussian and averaged across 9 repetitions of the illustrated leftwards reach (flanking traces show s.e.m.). **d**, Projection of the leech population response into the two-dimensional jPCA space. The two dimensions are plotted versus each other (top) and versus time (bottom). Units are arbitrary but fixed between axes. **e**, Similar projection for the walking monkey. **f**, Similar projection for the reaching monkey.

Fig. 1c, f are idiosyncratic. But, as shown below, rotations of the neural state are one of the most prominent features of the data.

Quasi-rhythmic responses during reaching

We analysed 469 single-neuron recordings from motor and premotor cortex of four monkeys (identified as A, B, J, N). We made a further 364 simultaneous recordings (single and multi-unit isolations) from two pairs of implanted 96-electrode arrays (monkeys J, N). Monkeys executed straight reaches (monkeys A, B) or straight and curved reaches (monkeys J, N). An instructed delay paradigm allowed monkeys to prepare their reaches before a go cue. We analysed 9 data sets, each using 27–108 reach types ('conditions'). For each neuron and condition we computed and analysed the average across-trial firing rate.

Most neurons exhibited preparatory and movement-related responses (Fig. 2). Responses were typically complex, multiphasic and heterogeneous¹⁴. Yet there appear to be oscillations in many single-neuron responses, beginning just before movement onset and lasting for ~1–1.5 cycles. These quasi-oscillatory patterns were seen for all reach types and all monkeys. Yet interpretational caution is warranted: multiphasic responses might exist for any number of reasons. The critical question is whether there exists orderly rotational structure, across conditions, at the population level.

We have proposed that motor cortex responses reflect the evolution of a neural dynamical system, starting at an initial state set by preparatory activity^{14,15,17,18,26}. If the rotations of the neural state (Fig. 1f) reflect straightforward dynamics, then similar rotations should be seen for all conditions. In particular, the neural state should rotate in the same direction for all conditions¹⁵, even when reaches are in opposition.

We projected the population response for all conditions onto the jPC plane. This was done for 200 ms of data, beginning when

preparatory activity transitions to movement-related activity (Supplementary Movie 3 shows a longer span of time). The resulting projections (Fig. 3a–f) show four notable features. First, rotations of

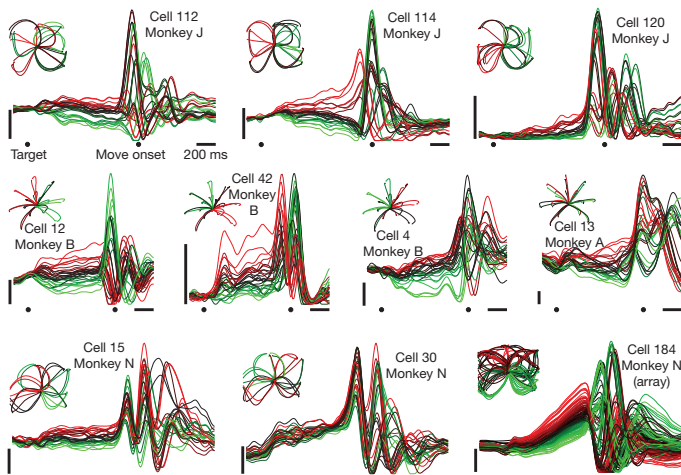


Figure 2 | Firing rate versus time for ten example neurons, highlighting the multiphasic response patterns. Each trace plots mean across-trial firing rate for one condition. Traces are coloured red to green based on the level of preparatory activity observed for that neuron. This allows inspection of how the pattern of preparatory tuning changes during the movement. Data were averaged separately locked to target onset, the go cue, and movement onset. To aid viewing, traces are interpolated across the gaps between epochs. Vertical scale bars indicate 20 spikes s⁻¹. Insets plot hand trajectories, which are different for each data set. Traces are coloured using the same code as for the neural data: red traces indicate those conditions with the greatest preparatory response.

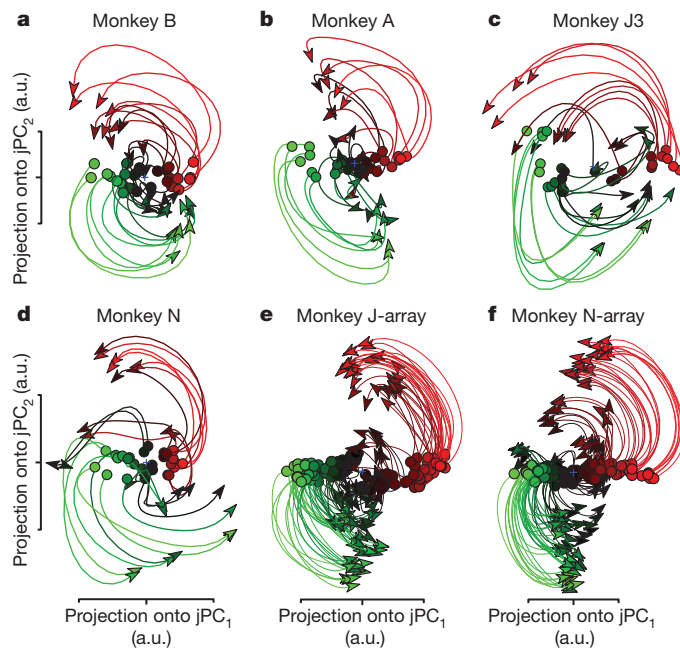


Figure 3 | Projections of the neural population response. **a**, Projection for monkey B (74 neurons; 28 straight-reach conditions). Each trace (one condition) plots the first 200 ms of movement-related activity away from the preparatory state (circles). Traces are coloured on the basis of the preparatory-state projection onto jPC_1 . a.u., arbitrary units. **b**, Projection for monkey A (64 neurons; 28 straight-reach conditions). **c**, Monkey J, data set 3 (55 neurons; 27 straight- and curved-reach conditions). **d**, Monkey N (118 neurons; 27 straight- and curved-reach conditions). **e**, Monkey J-array (146 isolations; 108 straight- and curved-reach conditions). **f**, Monkey N-array (218 isolations; 108 straight- and curved-reach conditions).

the neural state are prevalent during reaching. Second, the neural state rotates in the same direction across conditions (Fig. 3, different traces). Third, the rotation phase follows naturally from the preparatory state. Last, state-space rotations do not relate directly to reach curvature. Monkeys A and B executed straight reaches; monkeys J and N executed a mixture of straight reaches, clockwise-curving reaches, and anticlockwise-curving reaches (Fig. 2, insets). Yet for each data set the neural state rotates in the same direction across conditions. Rotations appear to reflect dynamics that are consistent across conditions, rather than the pattern of kinematics per se.

If the initial population-level preparatory state is known (Fig. 3, circles) subsequent states are reasonably predictable. Such predictability is absent at the individual-neuron level: the correlation between preparatory and movement tuning averages nearly zero¹⁵. Nevertheless, the ordered state-space rotations relate directly to the seemingly disordered single-neuron responses. Each axis of the jPCA projection captures a time-varying pattern that resembles a single-neuron response (Supplementary Fig. 1). Single neurons strongly reflect combinations of these underlying patterns. However, that underlying structure is not readily apparent when plotting each pattern alone, or each neuron individually¹⁵. Furthermore, the rotations during reaching are quasi-oscillatory, lasting only 1–1.5 cycles (Fig. 1c; also see Supplementary Movie 3). Their brevity and high frequency (up to ~ 2.5 Hz) makes them easy to miss unless trial counts are high (data sets averaged 810 trials per neuron) and data are precisely aligned on movement onset (Methods).

Controls regarding rotational structure

The central result of this study is the presence of the rotational patterns seen in Fig. 3. Those projections captured an average of 28% of the total data variance, and thus reveal patterns that are strongly present in the data. However, one must be concerned that

such patterns could have appeared by accident or for trivial reasons. To address this possibility, multiple ‘shuffle’ controls demonstrate that jPCA does not find rotational structure when such structure is not present (Supplementary Figs 2, 3 and Supplementary Movie 4). Similarly, rotations in the walking monkey were not erroneously found when the monkey was stationary (Supplementary Movie 1).

The fact that a single plane (two dimensions) captures an average of 28% of the total data variance is notable, given the high dimensionality of the data itself¹⁴. As a comparison, the dimensions defined by PC_2 and PC_3 (which by definition capture the second- and third-most data variance possible) together capture 29% of the total variance. Thus, the jPCA projection simply captures response patterns that were always present in the top PCs, but were difficult to see because they were not axis aligned. In fact, there were typically two or three orthogonal planes that captured rotational structure at different frequencies (Supplementary Fig. 4). Together these captured 50–70% of the total data variance. Thus, rotations are a dominant feature of the population response. This was true for primary motor cortex and dorsal premotor cortex independently (Supplementary Fig. 5).

Rotations, kinematics and EMG

Traditional views posit that motor cortex neurons are tuned for movement parameters such as direction. This perspective does not naturally account for the data in Fig. 3. We simulated neural populations that were directionally tuned for velocity with an additional non-directional sensitivity to speed²⁷. Simulated preparatory activity was tuned for reach direction and distance²⁸. We simulated one ‘velocity model’ data set per recorded data set, based upon the recorded velocities and endpoints. Firing rates, trial counts, neuron counts and spiking noise were matched to the recorded data. For velocity-model populations, jPCA found no robust or consistent rotations (Fig. 4a, d, h). This was true for all data sets (summary analysis below) including those with curved reaches (for example, Fig. 4h). We also simulated a ‘complex-kinematic’ model in which responses reflected the weighted sum of kinematic parameters (position, velocity, acceleration and jerk) that correlate with muscle activity⁶. This model produced multiphasic responses but not consistent rotations (Fig. 4b, e, i). We also recorded EMG (electromyograms) from a population of muscles (6–12 recordings per data set). Although EMG is strongly multiphasic, the population of muscles did not show consistent rotations (Fig. 4c, f, j; summary data below). This was not due to the smaller size of the muscle population (Supplementary Fig. 6). In sum, rotations in state space require more than multiphasic responses: they require a pair of multiphasic patterns with phases consistently $\sim 90^\circ$ apart. The neural population contains that complementary pair; the simulated and muscle populations do not. Still, EMG may bear some relationship to the observed rotations—a possibility explored below.

The rotations of the neural state are a robust feature of the physiological data, but it is not immediately apparent how those rotations relate to the reaches themselves. This question is especially relevant because the reaches were not overtly rhythmic. A possible answer is that muscle activity might be constructed from an oscillatory basis. To test whether this is plausible, we simulated a simple dynamical model possessing two orthogonal rotations in state space: one at a high frequency and one at a low frequency. Muscle activity was fit as the sum of the resulting oscillations in the temporal domain. For example, when fitting the deltoid EMG for dataset J3 (the third data set from monkey J) the higher-frequency rotation in the model occurred at 2.8 Hz (Fig. 5a). Different conditions (9 and 25 are shown) involved different amplitudes and phases, set by the preparatory state of the model. The vertical, ‘lagging’ dimension drove simulated muscle activity, and the projections onto that dimension (Fig. 5b, c, dark blue) provided key features of the EMG fit. The slower features are provided by a 0.3 Hz oscillation (not shown).

This ‘generator model’ provided excellent EMG fits (Fig. 5b, c and Supplementary Figs 7 and 8). The fit/EMG correlation ranged from

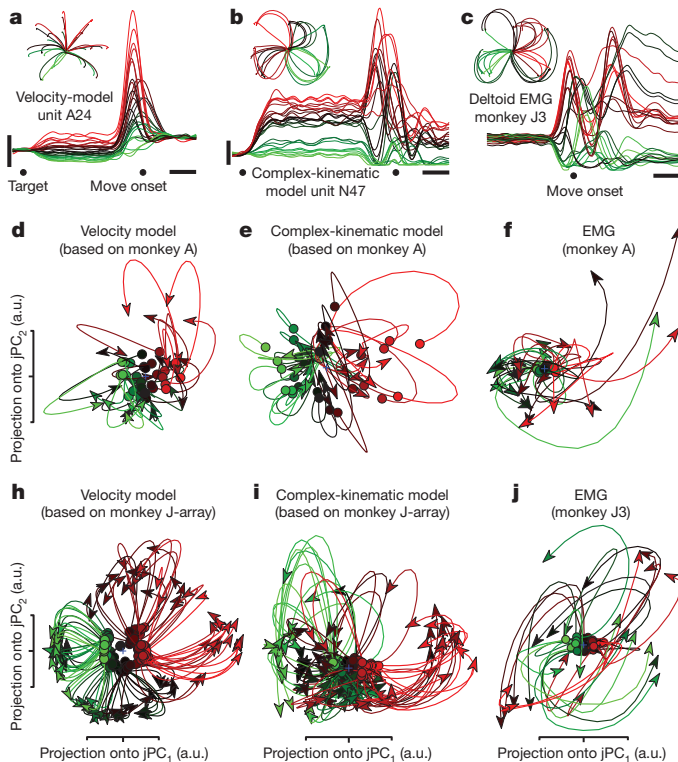


Figure 4 | Projections of simulated neural and muscle populations.

a. Simulated velocity-model unit, based on hand velocities of monkey A. The preferred direction points up and right. Presentation and scale bars as in Fig. 2. **b.** Simulated unit from the complex-kinematic model. **c.** EMG from the deltoid (monkey J, data set 3), coloured by initial response. EMG was rectified, smoothed, and averaged across trials. **d.** Projection of the velocity-model population response (64 simulated neurons) for monkey A. Identical presentation and analysis as Fig. 3. a.u., arbitrary units. **e.** Projection of the complex-kinematic model population response for monkey A. **f.** Projection of the recorded muscle population response for monkey A. **g.** Same as **d** but for monkey J-array. **h.** Same as **e** but for monkey J-array. **i.** Same as **f** but for monkey J, data set 3.

0.97–0.99 across data sets. Thus, deltoid activity could always be generated from the sum of two underlying rotations whose phases and amplitudes (but not frequencies) vary across conditions. This raises a subtle but key point: although EMG responses do not themselves exhibit state-space rotations, EMG can nevertheless be constructed from underlying rotations. This observation need not imply that cortex directly controls muscles. Yet it illustrates the plausibility of direct control⁶, and demonstrates that rotations provide a natural basis for generating non-rhythmic movement.

Faster reaches might have been expected to involve faster rotations, or longer reaches to involve longer rotations. However, EMG could be fit using two rotations with fixed frequency and duration. This was true even though the 27 conditions differed greatly in reach speed and duration (Fig. 5g). We return to this point below.

For representational models, individual-unit responses reflect the ‘factors’ for which those units are tuned. For a dynamical model, individual-unit responses should reflect the underlying dynamical factors: the patterns present on each axis of the state space. We simulated a population of generator-model units whose rates depended, with random weights, on these underlying patterns. ‘Preparatory’ activity was simply the initial state. Simulated units displayed multiphasic response patterns resembling those of real neurons (Fig. 5d, e and Supplementary Fig. 8). Both real and simulated responses exhibited reversals in ‘preferred condition’¹⁴ and a weak correlation between preparatory and movement-period ‘tuning’^{15,29}. Despite such surface complexity, jPCA projections of the simulated populations

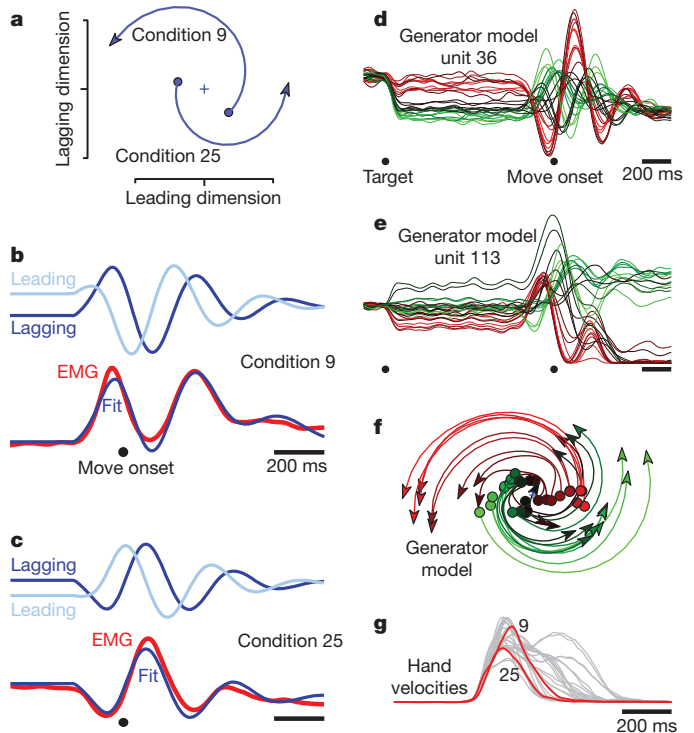


Figure 5 | Illustration of how a simple model generates fits to EMG using a pair of brief rotations. **a.** The higher-frequency rotation (2.8 Hz) is plotted for conditions 9 and 25, and shows the first 200 ms of evolution away from the preparatory state (filled circles). The preparatory state determines rotation amplitude and phase. One state dimension (‘leading’, on the x-axis) is always 90° ahead of the other (‘lagging’, on the y-axis). **b.** Projections onto the leading and lagging dimensions (light and dark blue) versus time (condition 9). The fit to the EMG is the sum of lagging components from the 2.8 Hz rotation (shown) and a lower frequency rotation (0.3 Hz, not shown). **c.** Similar to **b**, but for condition 25. The 2.8 Hz rotation is ~180° out of phase with that in **b**. **d.** Simulated response of a unit from the generator model (Methods). Presentation as in Fig. 2. **e.** A second simulated unit. **f.** jPCA projection of the simulated population; compare with the neural data in Fig. 3c. **g.** Hand velocities for the 27 conditions in **d**, **e**, **f**. Red traces show conditions 9 and 25.

successfully reveal the simple underlying rotations (Fig. 5f). These rotations resemble those observed for the neural data (Fig. 3).

Population-level quantification

To quantify rotation strength we measured the angle from the neural state in the jPCA plane (x) to its derivative (\dot{x}) for every condition and time (Fig. 6a). The first jPCA plane is oriented such that the average rotational tendency of the data—however weak or strong—is anti-clockwise. Thus, angles near positive $\pi/2$ indicate a strong rotational component. The neural data and the generator model have distributions with peaks near $\pi/2$. In contrast, the velocity-tuned model, the complex-kinematic model, and EMG all had distributions that peaked slightly above zero.

Rotation strength was also quantified via the jPCA computation, in which the data were fit with:

$$\dot{x}(t, c) = M_{\text{skew}} x(t, c) \quad (3)$$

where $x(t, c)$ is the k -dimensional ($k = 6$ for all figures) population state for time t and condition c . M_{skew} is a skew-symmetric matrix ($M_{\text{skew}} = -M_{\text{skew}}^T$) that captures rotational dynamics. The first two jPCs were the two eigenvectors associated with the largest-magnitude eigenvalues of M_{skew} , which indicate the strongest rotational plane in the dynamical system fit by equation (3). Projecting data onto those jPCs (as in Figs 3 and 4) reveals the prevalence of rotations. We can also assess rotation prevalence by quantifying how well M_{skew} fits the data relative to an unconstrained matrix M . That unconstrained

matrix M provides the best performance of any matrix (skew-symmetric matrices are a subset of unconstrained matrices). For the neural data and generator model, the fit provided by M_{skew} was nearly as good as the fit provided by the best unconstrained M (Fig. 6b). For the velocity-tuned and complex-kinematic models, M_{skew} performed poorly. Results were similar whether we considered all six dimensions or just the plane with the strongest rotations (the plane defined by the first two jPCs). Thus, of those dynamics that can be captured linearly, rotational dynamics dominate only for the generator model and neural data.

EMG data showed weak rotations (Fig. 6a, b, red), underscoring a central point: state-space rotations result not from a multiphasic

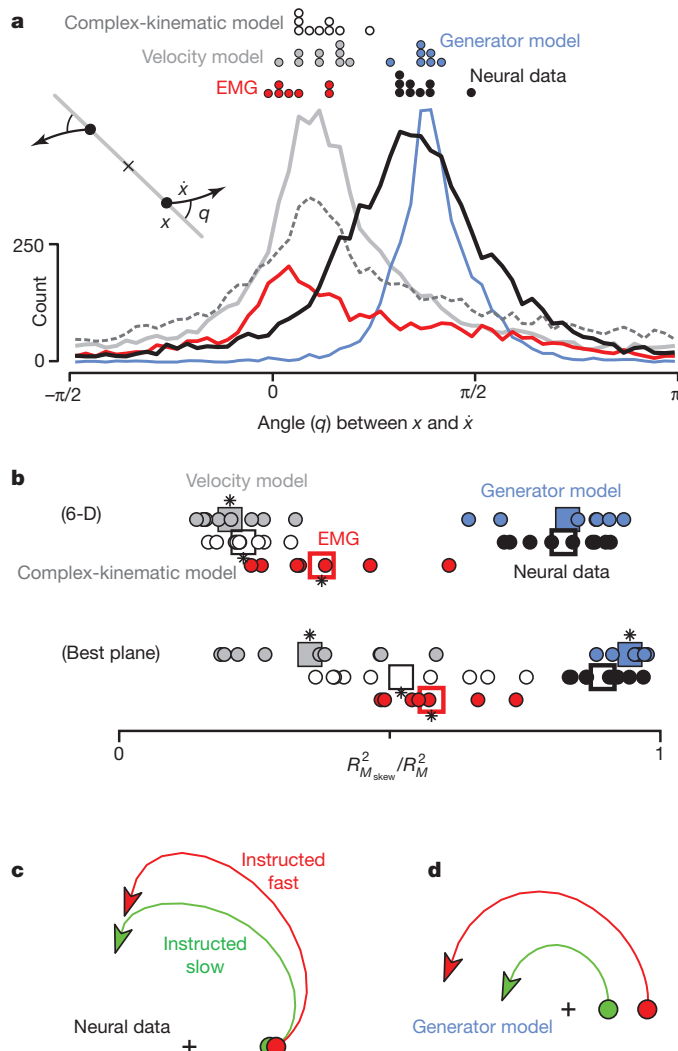


Figure 6 | Consistency of rotational dynamics for real and simulated data.

a, Histograms of the angle between the neural state, x , and its derivative, \dot{x} for real and simulated data. The angle was measured as illustrated schematically (inset) after projecting the data onto the first jPCA plane. Pure rotation results in angles near $\pi/2$; pure scaling/expansion results in angles near 0. Distributions include all analysed times and conditions. Dots at top show distribution peaks for individual data sets. **b**, Quality of the fit (R^2) provided by M_{skew} relative to an unconstrained M . We assessed fit quality for both the rank 6 matrices that capture dynamics in all 6 analysed dimensions (6-D; top row) and the rank 2 matrices that capture dynamics in the first (best) jPCA plane (bottom row). Circles plot performance for individual data sets. Squares give overall averages. Asterisks indicate a significant difference (t -test, $P < 0.05$) from neural data. **c**, Average (across monkeys A and B) neural trajectory for all instructed-slow conditions (green) and all instructed-fast conditions (red). **d**, Similar to c but for the generator model.

signal, but from how that signal is constructed. For example, the generator model exhibits rotations even though the EMG does not. More generally, many features of the observed rotations make sense in terms of how outputs (EMG, kinematics) might be generated, rather than in terms of the outputs themselves. For example, a strong intuition, and a prediction of most hypotheses of motor cortex, is that neural responses should unfold faster for faster movements. However, the generator model makes the opposite prediction; rotation frequencies are fixed. We tested this prediction using two data sets where target colour instructed fast versus slow reaches. Both the generator model and the neural data exhibited rotations that were of similar angular velocity for fast and slow reaches (Fig. 6c, d). The same point can be made by separately applying jPCA for fast and slow reaches: the largest eigenvalues of M_{skew} were actually slightly smaller for fast reaches (8.8 versus 9.8 radians s^{-1} for monkey A, 12.2 versus 13.5 radians s^{-1} for monkey B). Rotation amplitude, rather than frequency, differed between speeds (Fig. 6c, d). This finding is readily interpretable in light of the generator model: larger-amplitude rotations produce more strongly multiphasic responses, a feature of the EMG necessary to drive larger accelerations/decelerations (also see Supplementary Fig. 9).

Discussion

Rotations of the population state are a prominent feature of the cortical response during reaching. Rotations follow naturally from the preparatory state and are consistent in direction and angular velocity across the different reaches that each monkey performed. The rotational structure is much stronger and more consistent than expected from chance or previous models. These population-level rotations are a relatively simple dynamical feature yet explain seemingly complex features of individual-neuron responses, including frequent reversals of preferred direction^{14,30}, and the lack of correlation between preparatory and movement-period tuning^{15,29} (Fig. 5d, e and Supplementary Fig. 8). State-space rotations produce briefly oscillatory temporal patterns that provide an effective basis for producing multiphasic muscle activity, suggesting that non-periodic movements may be generated via neural mechanisms resembling those that generate rhythmic movement^{20,22,31–33}.

Recent results suggest that preparatory activity sets the initial state of a dynamical system, whose subsequent evolution produces movement activity¹⁵. Aspects of these dynamics—a rotation away from the preparatory state—appear straightforward. However, the circuitry that creates these dynamics is unclear; it may be purely local, or may involve recurrent circuitry³⁴ that links motor cortex with the spinal cord and with subcortical structures³⁵. Peripheral feedback is also expected to shape neural dynamics³⁶, although this cannot account for the first ~ 150 ms of ‘neural motion’ (the hand has yet to move). The finding that dynamics are captured by a skew-symmetric matrix suggests functionally antisymmetric connectivity: a given neural dimension (for example, jPC₁) positively influences another (for example, jPC₂), which negatively influences the first. However, it is unclear whether this population-level pattern directly reflects a circuit-level dominance of antisymmetric connectivity. We also stress that although rotations are a dominant pattern in the data across multiple dimensions (Supplementary Fig. 4), non-rotational components exist as well, perhaps reflecting the nonlinear dynamics necessary for initiating or terminating movement, for stability³⁷, and for feedback control^{16,38}.

It is hoped that a focus on the dynamics that generate movement will help transcend the controversy over what single neurons in motor cortex ‘code’ or ‘represent.’ Many of the neural response features that seem most baffling from a representational perspective are natural and straightforward from a population-level dynamical systems perspective. It therefore seems increasingly likely that motor cortex can be understood in relatively straightforward terms: as an engine of movement that uses lawful dynamics.

METHODS SUMMARY

Optical recordings from the isolated leech central nervous system were made by K. Briggman and W. Kristan and have been described previously^{24,25}. We recorded neural activity from trained monkeys using both single- and multi-electrode techniques. We recorded from the arm representation of premotor cortex using a wireless system while the monkey walked to obtain juice from the front of a treadmill. We recorded from the arm representation of motor and premotor cortex while monkeys reached to targets projected onto a vertically oriented screen, also for a juice reward. All surgical and animal care procedures were performed in accordance with National Institutes of Health guidelines and were approved by the Stanford University Institutional Animal Care and Use Committee.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 6 May 2011; accepted 5 April 2012.

Published online 3 June 2012.

- Lemon, R. N. An enduring map of the motor cortex. *Exp. Physiol.* **93**, 798–802 (2008).
- Evarts, E. V. Relation of pyramidal tract activity to force exerted during voluntary movement. *J. Neurophysiol.* **31**, 14–27 (1968).
- Mussa-Ivaldi, F. A. Do neurons in the motor cortex encode movement direction? An alternative hypothesis. *Neurosci. Lett.* **91**, 106–111 (1988).
- Fetz, E. E. Are movement parameters recognizably coded in the activity of single neurons? *Behav. Brain Sci.* **15**, 679–690 (1992).
- Sanger, T. D. Theoretical considerations for the analysis of population coding in motor cortex. *Neural Comput.* **6**, 29–37 (1994).
- Todorov, E. Direct cortical control of muscle activation in voluntary arm movements: a model. *Nature Neurosci.* **3**, 391–398 (2000).
- Hatsopoulos, N. G. Encoding in the motor cortex: was Evarts right after all? Focus on “motor cortex neural correlates of output kinematics and kinetics during isometric-force and arm-reaching tasks”. *J. Neurophysiol.* **94**, 2261–2262 (2005).
- Scott, S. H. Inconvenient truths about neural processing in primary motor cortex. *J. Physiol. (Lond.)* **586**, 1217–1224 (2008).
- Morrow, M. M., Pohlmeier, E. A. & Miller, L. E. Control of muscle synergies by cortical ensembles. *Adv. Exp. Med. Biol.* **629**, 179–199 (2009).
- Aflalo, T. N. & Graziano, M. S. A. Relationship between unconstrained arm movements and single-neuron firing in the macaque motor cortex. *J. Neurosci.* **27**, 2760–2780 (2007).
- Kalaska, J. F. From intention to action: motor cortex and the control of reaching movements. *Adv. Exp. Med. Biol.* **629**, 139–178 (2009).
- Georgopoulos, A. P., Schwartz, A. B. & Kettner, R. E. Neuronal population coding of movement direction. *Science* **233**, 1416–1419 (1986).
- Kakei, S., Hoffman, D. S. & Strick, P. L. Muscle and movement representations in the primary motor cortex. *Science* **285**, 2136–2139 (1999).
- Churchland, M. M. & Shenoy, K. V. Temporal complexity and heterogeneity of single-neuron activity in premotor and motor cortex. *J. Neurophysiol.* **97**, 4235–4257 (2007).
- Churchland, M. M., Cunningham, J. P., Kaufman, M. T., Ryu, S. I. & Shenoy, K. V. Cortical preparatory activity: representation of movement or first cog in a dynamical machine? *Neuron* **68**, 387–400 (2010).
- Todorov, E. & Jordan, M. I. Optimal feedback control as a theory of motor coordination. *Nature Neurosci.* **5**, 1226–1235 (2002).
- Graziano, M. S. New insights into motor cortex. *Neuron* **71**, 387–388 (2011).
- Churchland, M. M., Santhanam, G. & Shenoy, K. V. Preparatory activity in premotor and motor cortex reflects the speed of the upcoming reach. *J. Neurophysiol.* **96**, 3130–3146 (2006).
- Cisek, P. Preparing for speed. Focus on: “preparatory activity in premotor and motor cortex reflects the speed of the upcoming reach”. *J. Neurophysiol.* **96**, 2842–2843 (2006).
- Georgopoulos, A. P. & Grillner, S. Visuomotor coordination in reaching and locomotion. *Science* **245**, 1209–1210 (1989).
- Grillner, S. Biological pattern generation: the cellular and computational logic of networks in motion. *Neuron* **52**, 751–766 (2006).
- Rokni, U. & Sompolinsky, H. How the brain generates movement. *Neural Comput.* **24**, 289–331 (2011).
- Kristan, W. B. Jr & Calabrese, R. L. Rhythmic swimming activity in neurones of the isolated nerve cord of the leech. *J. Exp. Biol.* **65**, 643–668 (1976).
- Briggman, K. L. & Kristan, W. B. Jr. Imaging dedicated and multifunctional neural circuits generating distinct behaviors. *J. Neurosci.* **26**, 10925–10933 (2006).
- Briggman, K. L., Abarbanel, H. D. & Kristan, W. B. Jr. Optical imaging of neuronal populations during decision-making. *Science* **307**, 896–901 (2005).
- Shenoy, K. V., Kaufman, M. T., Sahani, M. & Churchland, M. M. in *Progress in Brain Research: Enhancing Performance for Action and Perception* (eds Green, A., Chapman, E., Kalaska, J. F. & Lepore, F.) (Elsevier, 2011).
- Moran, D. W. & Schwartz, A. B. Motor cortical representation of speed and direction during reaching. *J. Neurophysiol.* **82**, 2676–2692 (1999).
- Messier, J. & Kalaska, J. F. Covariation of primate dorsal premotor cell activity with direction and amplitude during a memorized-delay reaching task. *J. Neurophysiol.* **84**, 152–165 (2000).
- Kaufman, M. T. et al. Roles of monkey premotor neuron classes in movement preparation and execution. *J. Neurophysiol.* **104**, 799–810 (2010).
- Sergio, L. E., Hamel-Paquet, C. & Kalaska, J. F. Motor cortex neural correlates of output kinematics and kinetics during isometric-force and arm-reaching tasks. *J. Neurophysiol.* **94**, 2353–2378 (2005).
- Llinas, R. *I of the Vortex* (MIT Press, 2002).
- Yuste, R., MacLean, J. N., Smith, J. & Lansner, A. The cortex as a central pattern generator. *Nature Rev. Neurosci.* **6**, 477–483 (2005).
- Yakovenko, S., Krouchev, N. & Drew, T. Sequential activation of motor cortical neurons contributes to intralimb coordination during reaching in the cat by modulating muscle synergies. *J. Neurophysiol.* **105**, 388–409 (2011).
- Middleton, F. A. & Strick, P. L. Basal ganglia and cerebellar loops: motor and cognitive circuits. *Brain Res. Brain Res. Rev.* **31**, 236–250 (2000).
- Thorn, C. A., Atallah, H., Howe, M. & Graybiel, A. M. Differential dynamics of activity changes in dorsolateral and dorsomedial striatal loops during learning. *Neuron* **66**, 781–795 (2010).
- Herter, T. M., Korb, T. & Scott, S. H. Comparison of neural responses in primary motor cortex to transient and continuous loads during posture. *J. Neurophysiol.* **101**, 150–163 (2008).
- Sussillo, D. & Abbott, L. F. Generating coherent patterns of activity from chaotic neural networks. *Neuron* **63**, 544–557 (2009).
- Scott, S. H. Optimal feedback control and the neural basis of volitional motor control. *Nature Rev. Neurosci.* **5**, 532–546 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are deeply grateful to K. Briggman and W. Kristan for providing data recorded from the leech. We thank M. Risch for animal care, S. Eissensee for administrative support, and D. Haven and B. Oskotsky for information technology support. We thank Z. Ghahramani and C. Rasmussen for discussion of jPCA and related methods. We thank D. Sussillo, S. Grossman and M. Sahani for analysis suggestions and commentary on the manuscript. This work was supported by a Helen Hay Whitney postdoctoral fellowship and National Institutes of Health (NIH) postdoctoral training fellowship (M.M.C.), the Burroughs Wellcome Fund Career Awards in the Biomedical Sciences (M.M.C., K.V.S.), Engineering and Physical Sciences Research Council grant EP/H019472/1 and the McDonnell Center (J.P.C.), a National Science Foundation graduate research fellowship (M.T.K.), a Texas Instruments Stanford Graduate Fellowship (J.D.F.), a Paul and Daisy Soros Fellowship (P.N.), the Stanford Medical Scientist Training Program (P.N.), and these awards to K.V.S.: NIH Director's Pioneer Award (1DP1OD006409), NIH NINDS EUREKA Award (R01-NS066311), NIH NINDS BRP (R01-NS064318), NIH NINDS CRCNS (R01-NS054283), DARPA-DSO REPAIR (N66001-10-C-2010), Stanford Center for Integrated Systems, NSF Center for Neuromorphic Systems Engineering at Caltech, Office of Naval Research, and the Whitaker Foundation, the McKnight Foundation, the Sloan Foundation and the Weston Havens Foundation.

Author Contributions The jPCA method was designed by J.P.C. and M.M.C. M.M.C. and M.T.K. collected data from the reaching monkeys. J.D.F. and P.N. collected data from the walking monkey. S.I.R. led the array implantation surgeries. K.V.S. contributed to all aspects of the work. All authors discussed the results and commented on the analyses and manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to M.M.C. (mc3502@columbia.edu).

METHODS

Recordings and task design. Recordings from the isolated leech central nervous system were made by K. Briggman and W. Kristan and have been described previously^{24,25}. Recordings from monkey cortex were made using both a delayed reach task (with head restraint) and from an unrestrained monkey walking on a treadmill^{39–41}. Animal protocols were approved by the Stanford University Institutional Animal Care and Use Committee.

Most analyses concerned data collected during delayed reach tasks, for which our basic methods have been described previously^{15,29,42}. Briefly, four male rhesus monkeys (A, B, J and N) performed delayed reaches on a fronto-parallel screen. Delays ranged from 0–1,000 ms (the exact range varied by monkey). Only trials with delays >400 ms were analysed. Fixation was enforced (at the central spot) during the delay for monkeys J and N. We used two variants of a centre-out reaching task. In the ‘speed task’ monkeys A and B reached to radially arranged targets at two distances. Reach speed was instructed by target colour (28 total conditions)¹⁸. In the ‘maze task’ monkeys J and N made both straight reaches and reaches that curved around one or more intervening barriers. This task was beneficial because of the large variety of different reaches that could be evoked. Typically we used 27 conditions: each providing a particular arrangement of target and barriers. Monkey J performed the task for four different sets of 27 conditions, resulting in four data sets (J1–J4). For the monkey J-array and N-array data sets, 108 conditions were presented in the same recording session. Recordings from three of the four monkeys (A, B and J) have been analysed in prior publications (for example, ref. 15).

Recordings were made from primary motor cortex (M1, both surface and sulcal) and from the adjacent (caudal) aspect of dorsal premotor cortex (PMd). Supplementary Fig. 5 shows the central analysis divided by area. Seven of the nine data sets (monkey A, B, J1, J2, J3, J4 and N) were recorded with conventional single-electrode techniques. These data sets involved a total of 469 single-unit isolations. The other two data sets (monkey J-array, recorded 18 September 2009; monkey N-array, recorded 23 September 2010) used pairs of chronically implanted 96-electrode arrays (Blackrock Microsystems). These array-based data sets involved, respectively, 146 and 218 isolations (each a mix of single and multi-unit isolations).

Arrays were implanted after directly visualizing sulcal landmarks. Single-electrode recordings were guided by stereotaxic criteria, the known response properties of M1 and PMd, and the effects of microstimulation. For all monkeys, at some point the dura was reflected and the sulcal landmarks directly visualized. Recordings were medial to the arcuate spur and lateral to the precentral dimple. Recordings were not made within rostral PMd, near the arcuate sulcus. Sulcal M1, surface M1, and caudal PMd are contiguous. Although there are important differences in their average response properties (for example, delay period activity is more common in PMd), these differences are far from absolute: M1-like neurons are frequently found in caudal PMd and vice versa. Most analyses thus considered all neurons without attempting to divide them on the basis of either anatomy or response properties (although see Supplementary Fig. 5).

For the freely walking monkey, data were recorded from an array implanted in the arm representation of PMd. The times of threshold crossings on 32 of the 96 channels were wirelessly transmitted using the HermesD system^{39,41}. Behavioural data were recorded using a commercially available video camera. Juice was dispensed at one end of the treadmill, providing incentive for the monkey to walk continuously.

EMG data were collected as described previously⁴². EMG records were rectified, smoothed and averaged before further analysis. A total of 61 recordings were made from six muscle groups: deltoid, biceps brachii, triceps brachii, trapezius, latissimus dorsi and pectoralis. Most data sets contained multiple recordings from each muscle (for example, one from each of the three heads of the deltoid). The total number of EMG recordings for some data sets was thus as high as 12. EMG was recorded for all data sets except those recorded using arrays.

Computing average firing rate as a function of time. Average trial counts were high (an average of 810 trials per neuron). To ensure response features were not lost to averaging, a concerted effort was made to compute the average firing rate only over trials with nearly identical reach trajectories. This was done by training to a high level of stereotyped behaviour, and by discarding the few trials for which behaviour was not tightly stereotyped. Average firing rates were further de-noised by filtering with a Gaussian (20 or 24 ms depending on the data set) and using a custom-developed smoothing method that discards idiosyncratic features that are both small and not shared across conditions (see supplementary figure 4 in ref. 15). This method improves the signal-to-noise ratio without over-smoothing in the temporal domain, which was important for preserving high-frequency features of the response. This step aids the visualization of single-neuron firing rates, but had essentially no effect on any of the population-level analyses

(Supplementary Fig. 10). EMG recordings and ‘recordings’ of simulated neurons were processed using all the same steps as for the neural data.

Because the delay period and reaction time were variable, firing rates were computed separately locked to target onset, the go cue, and movement onset. For presentation (where one wishes to follow a trace through different epochs) we interpolated over the gaps between the three epochs.

Fitting the generator model to EMG. For the generator model, we directly simulated two state-space rotations. The goal was to start not by simulating the responses of individual units, but by directly simulating the underlying structure of the population data in state space. The two simulated rotations produced patterns that were summed to fit the EMG for the deltoid. For example, the deltoid EMG for data set J3 was fit using a 2.8 Hz rotation and a 0.3 Hz rotation. Each rotation consisted of leading and lagging sinusoids windowed by a gamma function, with the initial state extended backwards in time to mimic preparatory activity (for example, Fig. 5a, b, c). The amplitude and phase of that rotation was different for every condition, to allow the model to fit the different EMG patterns recorded for each condition. Importantly, for a given data set the rotation always had the same frequency regardless of condition, with a rise and decay defined by the same windowing gamma function (for example, the 2.8 Hz rotation was always at 2.8 Hz and the 0.3 Hz rotation was always at 0.3 Hz). This mimics a dynamical system that is the same across conditions except for an initial state that determines phase and amplitude. EMG was fit as the sum of the lagging sinusoids, one for each of the two frequencies.

Optimization involved two levels. At the level of each individual condition, the amplitudes and phases that provided the best fit were found via regression. Regression exploited the fact that every possible amplitude/phase of a sinusoid can be constructed via a linear combination of a sine and cosine. This step is thus both fast and guaranteed to find the best fit. Regression involved an offset term, which could be different for each condition. At the level of the whole data set, we numerically optimized the two frequencies, the mean and shape parameter of the windowing gamma function, and the time when oscillations began. Optimization was started from many initial parameter choices and the best fit was chosen.

Each condition’s simulated EMG is simply the sum of two windowed sinusoids and a variable offset. However, the central idea of the generator model is that those sinusoids result from rotations in an internal state space. The generator model thus embodied five basic patterns: the pair of leading and lagging dimensions that make up each rotation plus the offset.

Simulated neural data. We produced two classes of simulated neural data sets. The first class (the velocity model and complex-kinematic model) was based on a traditional framework in which units were cosine tuned for kinematic factors. The second class was based on the generator model describe above, which emulates a simple dynamical system. For both classes of model, the firing rates of individual units were assumed to depend upon underlying factors. For the velocity-tuned model, movement-period activity was based upon three underlying factors: horizontal reach velocity, vertical reach velocity, and reach speed (for example, see ref. 27). Each unit thus had a preferred direction in velocity space. Preparatory activity was based upon three additional underlying factors: horizontal reach endpoint, vertical reach endpoint, and peak reach speed. For the complex-kinematic model, we assumed that because muscle activity reflects a variety of kinematic factors (position, velocity, acceleration, jerk) neural activity might share this property^{6,30}. As with the velocity model, each unit was cosine tuned with a preferred direction in physical space. Simulated activity depended on motion in that preferred direction with the following sensitivities: 25 (spikes s⁻¹) m⁻¹, 10 (spikes s⁻¹)/(m s⁻¹), 1 (spikes s⁻¹)/(m s⁻²), 0.05 (spikes s⁻¹)/(m s⁻³). These constants are taken from a published model⁶, but have been adjusted as follows. First, the sensitivity to position has been reduced by half, otherwise it tended to dominate to an unrealistic degree. Second, a sensitivity to jerk has been added. This makes for a more stringent control (it increases the multiphasic aspects of the simulated responses) and captures the expectation that cortical activity might be more phasic than muscle activity. Preparatory activity was sensitive to target endpoint.

For the generator model, the underlying factors were the oscillatory patterns captured by the underlying state space. These patterns defined both the movement period and (via the initial state for each pattern) preparatory activity. Also included as underlying factors were the two gamma functions that defined the oscillation envelopes. These factors were the same across all conditions, and were included to mimic the overall change in excitability that is presumed to cause the waxing and waning of oscillations. The inclusion of these un-tuned factors had a similar impact on the generator-model neurons as did the non-directional speed factor for the velocity-tuned model neurons. However, although the addition of these un-tuned factors served to increase the variety and realism of the simulated responses, it has essentially no impact on the main analyses (Figs 5g and 6). Those analyses are sensitive only to response aspects that differ among conditions.

The firing rate of each simulated unit was a random combination of the underlying factors. For the velocity and complex-kinematic models, the random combinations resulted in a roughly uniform range of preferred directions for both the preparatory and movement periods. For the generator model, the random combinations resulted in simulated units that typically reflected both oscillation frequencies, with a roughly uniform distribution of phases. This is indeed the default expectation for a large network that supports two oscillatory modes.

To produce simulated data with realistic levels of noise, we 'recorded' simulated spikes that were produced via a gamma-interval process (order 2) based on the underlying firing rate. For each neural data set, we simulated one unit for every recorded neuron, and matched the overall firing rates and trial counts of each simulated unit to those of the respective recorded neuron. The simulated spiking data was then analysed just as for the actual neural data. The velocity and complex-kinematic models each produced nine simulated data sets (one for every real data set). The generator model produced seven: it could not be simulated for the J-array or N-array data sets, as we did not attempt to record EMG for those 108-condition experiments.

Projections that capture rotational structure. We produced projections of the population data using a novel dimensionality reduction method, jPCA, designed for the present application. For most analyses we analysed 200 ms of time, sampled every 10 ms, starting just before the rapid change in neural activity that precedes movement onset. Before applying jPCA, a number of pre-processing steps were applied to the data (these same steps were also applied to the simulated data and EMG). Responses were normalized to have a similar firing-rate range for all neurons. 'Soft' normalization was used, so that neurons with very strong responses were reduced to approximately unity range, but neurons with weak responses had less than unity range. For each neuron, the data were mean-centred at every time: the average across-condition response was subtracted from the response for each condition. Thus, all subsequent analysis focused on those aspects of the neural response that differ across conditions. This pre-processing step can be skipped (see Supplementary Fig. 11) but the resulting projections often capture rotations that are similar for all conditions. In such cases one fails to gain multiple views of the underlying process, making it difficult to infer whether rotations are due to dynamics or to more trivial possibilities. It was thus deemed more conservative to only interpret projections where activity unfolds differently across conditions. Related population analyses (for example, the population vector) achieve the same end implicitly: non-directional aspects of the response cancel out. The pre-processing steps (and all subsequent analysis steps) were applied in the same way to all data sets, real and simulated.

The most critical pre-processing step was the use of traditional PCA. We compiled a data matrix, X , of size $n \times ct$, where n is the number of neurons, c is the number of conditions, and t is the number of time points. This matrix simply contains the firing rates of every neuron for every condition and every analysed time. We then used PCA to reduce the dimensionality of X from $n \times ct$ to $k \times ct$. $k = 6$ for all analyses in the main text, which is conservative given the true dimensionality of the data¹⁴. The resulting $6 \times ct$ matrix, X_{red} , defines a six-dimensional neural state for every time and condition. By pre-processing with PCA, we ensure that when jPCA is subsequently applied, it reveals only patterns of activity that are strongly present across neurons. Pre-processing with PCA greatly reduces any potential concern that the observed rotations were found

simply by looking in a very high-dimensional space (also see shuffle controls in Supplementary Figs 2 and 3).

jPCA is a method for finding projections (onto an orthonormal basis) that captures rotational structure in the data. jPCA is based on a comparison of the neural state with its derivative. We computed \dot{X}_{red} , of size $6 \times c(t-1)$ by taking the difference in the state between adjacent time points within each row of X_{red} . We then fit using

$$\dot{X}_{\text{red}} = M X_{\text{red}} \quad (4)$$

and

$$\dot{X}_{\text{red}} = M_{\text{skew}} X_{\text{red}} \quad (5)$$

(To keep the dimensions appropriate, the final time point for each condition in X_{red} was removed so that it was also $6 \times c(t-1)$). Thus, we are attempting to find matrices M and M_{skew} that take the state at each time in X_{red} , and transform it into the derivative of the state in \dot{X}_{red} . M can be found using linear regression. Finding M_{skew} requires more complex (but still linear) operations (see Supplementary Derivation for more detail). The quality of the above fits was assessed using R^2 (for example, Fig. 6b). R^2 captures the ability to linearly predict the data in \dot{X}_{red} (across all times and conditions) from the data in X_{red} .

M_{skew} has imaginary eigenvalues, and thus captures rotational dynamics. The strongest, most rapid rotations in the dynamical system occur in the plane defined by the eigenvectors associated with the largest two (complex-conjugate) imaginary eigenvalues. These eigenvectors (V_1 and V_2) are complex, but the associated real plane can be found by: $\text{jPC}_1 = V_1 + V_2$, and $\text{jPC}_2 = j(V_1 - V_2)$ (after resolving the ambiguity in the polarity of V_1 and V_2 such that their real components have the same sign). The first jPCA projection is then $X_{\text{jPCA}} = (\text{jPC}_1; \text{jPC}_2) \times X_{\text{red}}$. The matrix X_{jPCA} is thus of size $2 \times ct$, and describes the neural state, projected onto two dimensions, for every time and condition. For a given jPCA plane, the choice of orthogonal vectors (jPC_1 and jPC_2) within that plane is arbitrary. We therefore selected jPC_1 and jPC_2 so that any net rotation was anticlockwise (the same choice was of course used across all conditions for a given data set) and so that the spread of preparatory states was strongest along jPC_1 . We also computed the proportion of the total data variance captured by the jPCA plane, in a manner exactly analogous to that for PCA.

It is worth stressing that the six jPCs form an orthonormal basis that spans exactly the same space as the first six PCs. Thus, all patterns seen in the jPCA projections are also present in the PCA projections (the rotational patterns are simply not axis aligned in the latter case, and are thus less obvious to the eye; see Supplementary Movie 2).

39. Gilja, V., Chestek, C. A., Nuyujukian, P., Foster, J. D. & Shenoy, K. V. Autonomous head-mounted electrophysiology systems for freely behaving primates. *Curr. Opin. Neurobiol.* **20**, 676–686 (2010).
40. Foster, J. D. et al. in *Proc. of the 5th International IEEE EMBS Conference on Neural Engineering* 613–615 (IEEE, 2011).
41. Miranda, H. et al. A high-rate long-range wireless transmission system for simultaneous multichannel neural recording applications. *IEEE Trans. Biomed. Circ. Syst.* **4**, 181–191 (2010).
42. Churchland, M. M., Yu, B. M., Ryu, S. I., Santhanam, G. & Shenoy, K. V. Neural variability in premotor cortex provides a signature of motor preparation. *J. Neurosci.* **26**, 3697–3712 (2006).

Embryonic stem cell potency fluctuates with endogenous retrovirus activity

Todd S. Macfarlan^{1†}, Wesley D. Gifford¹, Shawn Driscoll¹, Karen Lettieri¹, Helen M. Rowe², Dario Bonanomi¹, Amy Firth³, Oded Singer³, Didier Trono² & Samuel L. Pfaff¹

Embryonic stem (ES) cells are derived from blastocyst-stage embryos and are thought to be functionally equivalent to the inner cell mass, which lacks the ability to produce all extraembryonic tissues. Here we identify a rare transient cell population within mouse ES and induced pluripotent stem (iPS) cell cultures that expresses high levels of transcripts found in two-cell (2C) embryos in which the blastomeres are totipotent. We genetically tagged these 2C-like ES cells and show that they lack the inner cell mass pluripotency proteins Oct4 (also known as Pou5f1), Sox2 and Nanog, and have acquired the ability to contribute to both embryonic and extraembryonic tissues. We show that nearly all ES cells cycle in and out of this privileged state, which is partially controlled by histone-modifying enzymes. Transcriptome sequencing and bioinformatic analyses showed that many 2C transcripts are initiated from long terminal repeats derived from endogenous retroviruses, suggesting this foreign sequence has helped to drive cell-fate regulation in placental mammals.

The zygote and its daughter cells are totipotent because they are able to develop into all embryonic and extraembryonic cell types^{1,2}. The progeny of these first two daughter cells become progressively more fate-restricted as they activate distinct patterns of gene expression that first direct them towards one of three broad lineages: Oct4⁺ Sox2⁺ Nanog⁺ epiblast cells that give rise to the embryo, Gata4⁺/6⁺ primitive endoderm cells that contribute to extraembryonic membranes that encase the embryo, and Cdx2⁺ trophoblast cells that form a large part of the placenta³. These early cell-fate decisions represent a major and relatively recent advance in mammalian evolution in which the placenta and extraembryonic tissues that support the intrauterine nourishment of the fetus allow development to progress further before birth. The epigenetic landscape of the zygote changes markedly during the first cell divisions. Shortly after fertilization, the oocyte maternal transcripts are replaced with newly synthesized RNAs generated by activating transcription of the zygotic genome^{4–6}. The unique transcriptional profile of the zygote and its daughter cells defines a brief period when the cells are totipotent.

Mouse ES cells are isolated from the inner cell mass (ICM) of blastocysts that have already become a separate lineage from the trophoblast^{7,8}. ICM-derived ES cells are regarded as pluripotent because they have the capacity to generate tissues of the fetus, but they are extremely inefficient at colonizing the extraembryonic tissues⁹. The rare contribution of ES cells to extraembryonic tissues could be explained by contamination of ES cultures with trophoblast or primitive endoderm-committed cells, or may occur because rare ES cells have acquired the ability to produce extraembryonic tissues in addition to embryonic tissues. This latter possibility is intriguing, because recent evidence shows that ES cell cultures are a heterogeneous mixture of metastable cells with fluctuating expression of genes such as *Zscan4*, *stella* (also known as *Dppa3*), *Nanog*, *Sox17* and *Gata6*, which could account for special attributes of individual cells^{10–14}.

A large number of retrotransposons are expressed when the zygotic genome is first transcribed, including the endogenous retroviruses

(ERVs), long interspersed nuclear element-1 (LINE-1), and the non-autonomous short interspersed elements (SINEs)¹⁵. At the 2C stage, murine endogenous retrovirus with leucine tRNA primer (MuERV-L, also known as MERVL and Erv4) elements are transiently derepressed and produce 3% of the transcribed messenger RNAs^{15–17}. After the 2C stage, MuERV-L retroelement expression is silenced^{18,19}. We discovered that this regulated pattern of MuERV-L expression overlapped with more than 100 2C-specific genes that have co-opted regulatory elements from these foreign retroviruses to initiate their transcription. We exploited the regulated activity of these 2C virus-derived promoters to label cells, and found that both ES and iPS cell cultures contain a small but relatively constant fraction of cells that has entered into the 2C-transcriptional state. Purification of these 2C-like cells shows that they have unique developmental characteristics and efficiently produce progeny for extraembryonic and embryonic lineages.

Identification of a 2C-like state within ES cultures

To identify zygotically activated genes we performed deep RNA sequencing (RNA-seq) on mouse oocytes and 2C-stage embryos. A comparison of the transcripts in these cells identified a large number of genes and retrotransposons that became expressed in the 2C embryo, as well as numerous transcripts that were downregulated (Fig. 1a and Supplementary Table 1). The most highly activated repeat was the MuERV-L family of retroviruses and their corresponding long terminal repeat (LTR) promoters (Mt2_mm), which were activated more than 300-fold (Supplementary Table 1). Sequence alignments showed that more than 25% of the nearly 700 copies of MuERV-L elements were activated, and that 307 genes generated chimaeric transcripts with junctions to MuERV-L elements (Fig. 1a and Supplementary Table 2), including 10 that were previously described¹⁵. Of the 626 chimaeric transcripts generated, >90% were 5' LTR–exon fusions that generated open reading frames (ORFs), suggesting that these LTRs had become functional promoters for

¹Howard Hughes Medical Institute, Gene Expression Laboratory, The Salk Institute for Biological Studies, 10010 North Torrey Pines, La Jolla, California 92037, USA. ²School of Life Sciences, Ecole Polytechnique Fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland. ³Laboratory of Genetics, The Salk Institute for Biological Studies, 10010 North Torrey Pines, La Jolla, California 92037, USA. [†]Present address: Program in Genomics of Differentiation, Eunice Kennedy Shriver National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, Maryland 20892, USA.

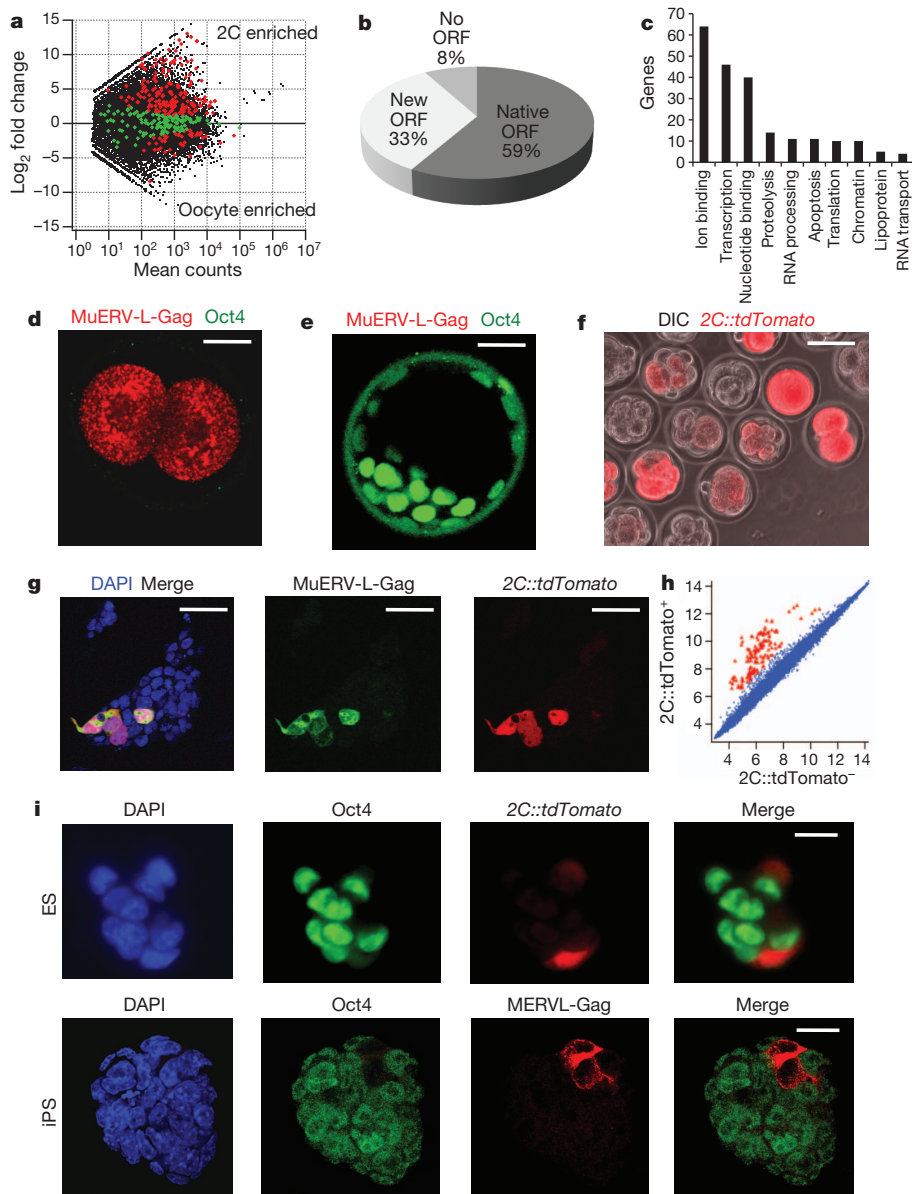


Figure 1 | The MuERV-L retrovirus and a reporter driven by its LTR marks the 2C state. **a**, Comparison of gene expression between oocytes and 2C embryos. Genes generating junctions to MuERV-L are shown in red and green, with those in red denoting significant change in expression. **b**, ORF status of predicted MuERV-L-linked chimaeric transcripts. **c**, Gene Ontology (GO) analysis of MuERV-L-linked protein-coding transcripts. The number of genes from the ten most enriched GO categories are shown. **d**, **e**, 2C (**d**) and blastocyst (**e**) embryos were mixed and immunostained with MuERV-L-Gag and Oct4

antibodies. Scale bars, 20 μ m. **f**, Zygotes were injected with the 2C::tdTomato transgene, and allowed to develop *in vitro* for 48 h before imaging. DIC, differential interference contrast. Scale bar, 50 μ m. **g**, 2C::tdTomato⁺ ES cells express MuERV-L-Gag protein, as detected by immunofluorescence. DAPI, 4',6-diamidino-2-phenylindole. Scale bars, 50 μ m. **h**, Microarray analysis of 2C::tdTomato⁺ and 2C::tdTomato⁻ cells. Red indicates genes with a greater than fourfold change in expression. **i**, 2C::tdTomato⁺ MuERV-L-Gag⁺ ES and iPS cells lack Oct4 protein, as determined by immunofluorescence. Scale bars, 20 μ m.

protein-coding genes (Fig. 1b and Supplementary Fig. 1a). The most significantly enriched Gene Ontology categories representing these chimaeric proteins were regulation of transcription, ion binding, translation, nucleotide binding and mRNA transport (Fig. 1c). Two notable transcription factors that used alternate MuERV-L-LTR promoters were *Gata4* and *Tead4*, which are important for the specification of primitive endoderm and trophoblast, respectively^{20–22}.

Because more than 300 of the nearly 700 copies of the MuERV-L endogenous retroviruses still encode Gag viral protein, we stained 2C and early blastocyst embryos to confirm that viral Gag was expressed and developmentally regulated. We found that 2C embryos express Gag but lack the pluripotency marker Oct4, whereas blastula cells lack Gag but express Oct4 (Fig. 1d, e). Thus, MuERV-L activity is developmentally regulated and these retroviral promoters have been co-opted by many cellular genes to impose tight control over their expression.

Next we asked whether it was possible to use the regulatory sequences from MuERV-L elements to label 2C cells. We cloned the MuERV-L 5' LTR, the primer-binding site, and a portion of the *gag* gene upstream of the red fluorescent protein tandem dimeric Tomato (tdTomato). We injected fertilized eggs with the 2C::tdTomato construct and monitored the expression of tdTomato during culture *in vitro*. tdTomato expression was highest in arrested zygotes and 2C embryos and became downregulated at the morula stage (Fig. 1f and Supplementary Movie 1). Notably, when we introduced the 2C::tdTomato construct into ES cells and selected for clonal stable integrants, we found several colonies that contained 1–5 cells that were strongly labelled with tdTomato among cells lacking expression of the reporter (Fig. 1g). We also found that rare ES cells expressed MuERV-L mRNA and Gag protein, and that these overlapped with 2C::tdTomato⁺ cells (Fig. 1g and Supplementary Fig. 1b, c). The

correspondence between the *2C::tdTomato* reporter and MuERV-L expression was further confirmed by immunoblotting, and electron microscopy imaging of viral epsilon particles encoded by MuERV-L within the endoplasmic reticulum of tdTomato⁺ cells but not tdTomato⁻ cells (Supplementary Fig. 1d, e). Thus, MuERV-L expression is restricted *in vivo* to 1–4-cell-stage embryos and is reactivated within a small subpopulation of ES cells derived from blastocysts.

To characterize the unexpected *2C::tdTomato*-labelled cells within ES cultures, we sorted tdTomato⁺ and tdTomato⁻ cells and performed microarray and mRNA sequencing analyses (Fig. 1h and Supplementary Tables 3–5). tdTomato⁺ cells expressed 55-fold higher levels of MuERV-L transcripts than tdTomato⁻ cells, but the vast majority of other retrotransposons were unaffected (Supplementary Table 3). Notably, tdTomato⁺ cells had 165 transcripts activated more than fourfold, and no genes repressed more than fourfold compared with tdTomato⁻ cells (Fig. 1h, Supplementary Table 4 and Supplementary Fig. 2a–f). Among the genes that were highly enriched in tdTomato⁺ cells, several were previously shown to be restricted to the 2–4-cell stage of development, including *Zscan4*, *Tcstv1/3*, *Eif1a*, *Gm4340* (also known as *Thoc4*), *Tdpz1-5* and *Zfp352* (refs 23–25). In total, 525 genes that were enriched in *2C::tdTomato*⁺ cells were also activated at the 2C stage, including 52 genes that generated chimaeric transcripts linked to MuERV-L elements (Supplementary Tables 6 and 7).

A hallmark of the ICM and ES cells is their expression of Oct4, Sox2 and Nanog, whereas totipotent 2C embryos do not express Oct4 (Fig. 1d, e). We found that *2C::tdTomato*⁺ cells within ES cultures also lacked Oct4, Sox2 and Nanog (Fig. 1i and Supplementary Fig. 1f). The reduction in Oct4, Sox2 and Nanog protein labelling occurred despite no changes in their mRNA levels, suggesting that the regulation is occurring post-transcriptionally (Fig. 1h–i and Supplementary Fig. 2g). In summary, *2C::tdTomato* labels a subset of ES cells that share transcriptional and proteomic features of 2C embryos and display markedly different patterns of pluripotency markers from most ES cells in culture.

ES cells cycle in and out of the 2C state

We considered the possibility that the expression of the *2C::tdTomato* reporter and MuERV-L-Gag protein in sporadic cells within ES

cultures might arise from contamination with trophectoderm or primitive endoderm. To exclude this possibility, we examined iPS cells derived from mouse fibroblasts because they should not be contaminated with cells from blastocyst embryos. Similar to ES cells, we found that sporadic iPS cells express the MuERV-L-Gag protein and lack Oct4 (Fig. 1i). Thus, the heterogeneity within ES cultures is a property that is shared with iPS cell cultures and is unlikely to arise from a cell contaminant.

Next we examined whether the *2C::tdTomato*⁺ cells represent a stable cell population or whether ES cells transition in and out of this 2C-like state. We used a Cre/loxP fate-mapping strategy to indelibly mark cells that had expressed 2C genes (Supplementary Fig. 3a–c). We generated a transgenic mouse line using the MuERV-L regulatory elements driving expression of a tamoxifen-inducible Cre recombinase (*2C::ERT2-Cre-ERT2*; Supplementary Fig. 3a). These mice were then mated with Cre-responsive reporter lines (*ROSA::LSL-tdTomato* and *ROSA::LSL-LacZ*; Supplementary Fig. 3b). ES cell lines were derived from double-positive transgenic blastocysts (Supplementary Fig. 3c). After addition of 4-hydroxytamoxifen (4HT) to the ES cultures we detected nuclear Cre expression in MuERV-L-Gag⁺ cells (Supplementary Fig. 3d). When ES cultures were grown for 2–6 days with 4HT we found a steady increase in the percentage of reporter-positive cells (Fig. 2a). Remarkably, over extended passages nearly every ES cell activated the reporter (Fig. 2b), demonstrating this transient state is regularly entered by ES cells.

To monitor the kinetics of the interconversion between *2C::tdTomato*⁺ and *2C::tdTomato*⁻ cells we performed flow cytometry to collect tdTomato⁺ and tdTomato⁻ cells. When these purified subpopulations were cultured we found that tdTomato⁺ cells produced tdTomato⁻ cells and vice versa (Fig. 2c). Within 24 h nearly 50% of the tdTomato⁺ cells convert to tdTomato⁻, independently of the starting percentages of the two different cell populations (Fig. 2c and data not shown). Under hypoxic conditions (5% O₂), the percentage of cells expressing the *2C::tdTomato* reporter was decreased, which could be reversed by shifting the cultures back to 20% O₂ (Fig. 2d). We also found that growing cells for 48 h in ‘ground-state’ media conditions (2i media²⁶) reduced but did not eliminate the presence of tdTomato⁺ cells relative to media containing knockout serum replacement, suggesting

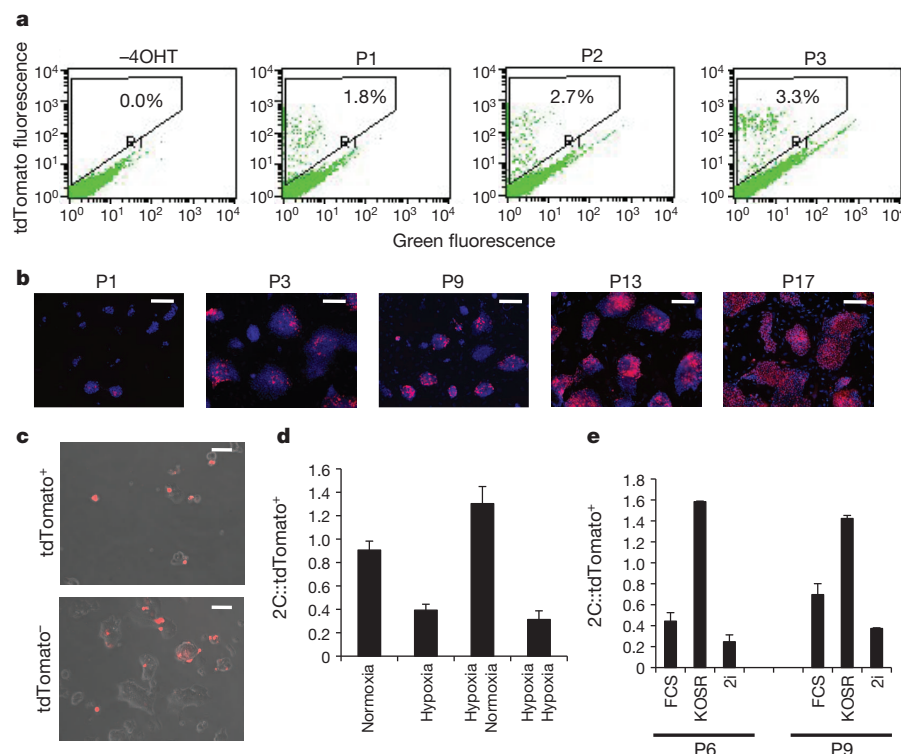


Figure 2 | ES cells enter the 2C state regularly, but remain in the state transiently owing to cell intrinsic and extrinsic factors. **a**, FACS analysis of *2C::ERT2-Cre-ERT2*, *ROSA::LSL-tdTomato* ES cells at increasing passage (P) in the presence of 4HT. The percentage of tdTomato⁺ cells is indicated. **b**, *2C::ERT2-Cre-ERT2*, *ROSA::LSL-LacZ* ES cells were cultured in the presence of 4HT, and at increasing passage, cells were fixed and immunostained with anti-β-galactosidase antibodies and counterstained with DAPI. Scale bars, 50 μm. **c**, *2C::tdTomato*⁺ and *2C::tdTomato*⁻ cells were collected by FACS and plated before imaging 48 h later. Scale bars, 50 μm. **d**, *2C::tdTomato* ES cells were cultured in 20% O₂ (normoxia) or 5% O₂ (hypoxia) for 48 h, or 48 h sequentially, and the percentage of tdTomato⁺ cells was determined by FACS. Error bars represent s.d., *n* = 3. **e**, *2C::tdTomato* ES cells at the indicated passage were cultured in media containing 15% fetal calf serum (FCS), 20% knockout serum replacement (KOSR) or N2B27 media containing 3 mM glycogen synthase kinase 3β (GSK3β) and mitogen-activated protein-kinase (MEK) inhibitors (2i) for 48 h before counting the percentage of *2C::tdTomato*⁺ cells by FACS. Error bars represent s.d., *n* = 3.

extrinsic and intrinsic mechanisms regulate the MuERV-L and 2C gene network (Fig. 2e).

The 2C-ES switch is regulated by histone modification

After activation of the zygotic genome in mouse development, histone deacetylation and histone H1 synthesis lead to the formation of repressive chromatin that is thought to limit the broad pattern of transcription present in 2C embryos^{27,28}. Using indirect immunofluorescence, we found that tdTomato⁺ cells had markedly higher levels of active histone marks, including methylation of histone 3 lysine 4 (H3K4) and acetylation of H3 and H4 (Supplementary Fig. 4a), a finding confirmed using immunoblot analysis of sorted cell populations (Fig. 3a). This type of chromatin mirrors that found in 2C embryos²⁸. Next we tested whether tdTomato⁺ cells had different levels of DNA methylation compared with non-labelled ES cells. We found that the MuERV-L sequences were hypomethylated in tdTomato⁺ cells compared with tdTomato⁻ cells. In contrast to the MuERV-L sequences, intracisternal A-type particle retroviruses were highly methylated in both tdTomato⁺ and tdTomato⁻ cells, suggesting the altered pattern of methylation was not uniform across the genome (Supplementary Fig. 4b). In summary, these data suggest that as ES cells (re)enter into the 2C state, their chromatin and DNA is altered to favour transcription in a way that mirrors the 2C embryo.

We previously demonstrated that MuERV-L and 2C-specific genes were increased in mutant ES cells lacking the histone lysine-specific demethylase gene *Kdm1a* (also known as *LSD1*)²⁹. To test whether other proviral co-repressors and histone-modifying enzymes also influence 2C-specific gene expression we profiled the transcriptome of ES cells with homozygous mutations in the KRAB (Kruppel-associated box)-associated transcriptional repressor *Kap1* and the H3K9 histone methyltransferase *G9a*^{29–31}. We found that MuERV-L and several 2C genes were significantly upregulated in *Kdm1a*, *Kap1* and *G9a* mutant ES cells (Fig. 3b, Supplementary Fig. 5b, c and

Supplementary Table 7). These findings were confirmed using *in situ* hybridization and immunofluorescence microscopy (Fig. 3c and Supplementary Fig. 5d). Treatment of 2C::tdTomato ES lines with the histone deacetylase inhibitor trichostatin A also increased the number of tdTomato⁺ cells fourfold (Fig. 3d). To understand better how 2C gene regulation is controlled when chromatin repressors are acutely eliminated we used a stably integrated 2C::tdTomato ES line that is homozygous for a floxed allele of *Kdm1a* and contains a *Cre-ERT* transgene that can be activated with 4HT. Within 24 h of deleting *Kdm1a* we found a tenfold increase in tdTomato⁺ cells that was steadily maintained (Fig. 3e). In addition, fluorescence-activated cell sorting (FACS)-purified tdTomato⁻ cells more rapidly became tdTomato⁺ in the absence of *Kdm1a* (Fig. 3f), and stayed in this state longer (Supplementary Movies 2 and 3). These findings suggest that *Kdm1a*, *Kap1*, *G9a* and histone deacetylases all contribute to the repression of 2C genes in ES cells, and that they function by altering the equilibrium between the 2C::tdTomato⁺ and 2C::tdTomato⁻ states.

2C-like ES cells have expanded fate potential

Because 2C-like cells within ES cultures express high levels of 2C-restricted genes found in totipotent blastomeres and reduced levels of pluripotency-associated proteins, we reasoned that this subpopulation of ES cells might have distinct functional characteristics. We tested whether 2C::tdTomato⁺ cells have acquired the ability to produce extraembryonic tissues, a characteristic that ES cells lack. We used FACS to collect tdTomato⁺ and tdTomato⁻ cells from a 2C::tdTomato ES line, and injected four cells into morula-stage embryos. The tdTomato⁻ cells contributed exclusively to the ICM of all five chimaeric blastocysts analysed (Fig. 4a). By contrast, the tdTomato⁺ cells contributed to the trophectoderm (in four out of five chimaeric embryos) in addition to the ICM (in three out of five chimaeric embryos) (Fig. 4a). To track the fate of the 2C::tdTomato ES cells later in development, we injected blastocysts with tdTomato⁺ or

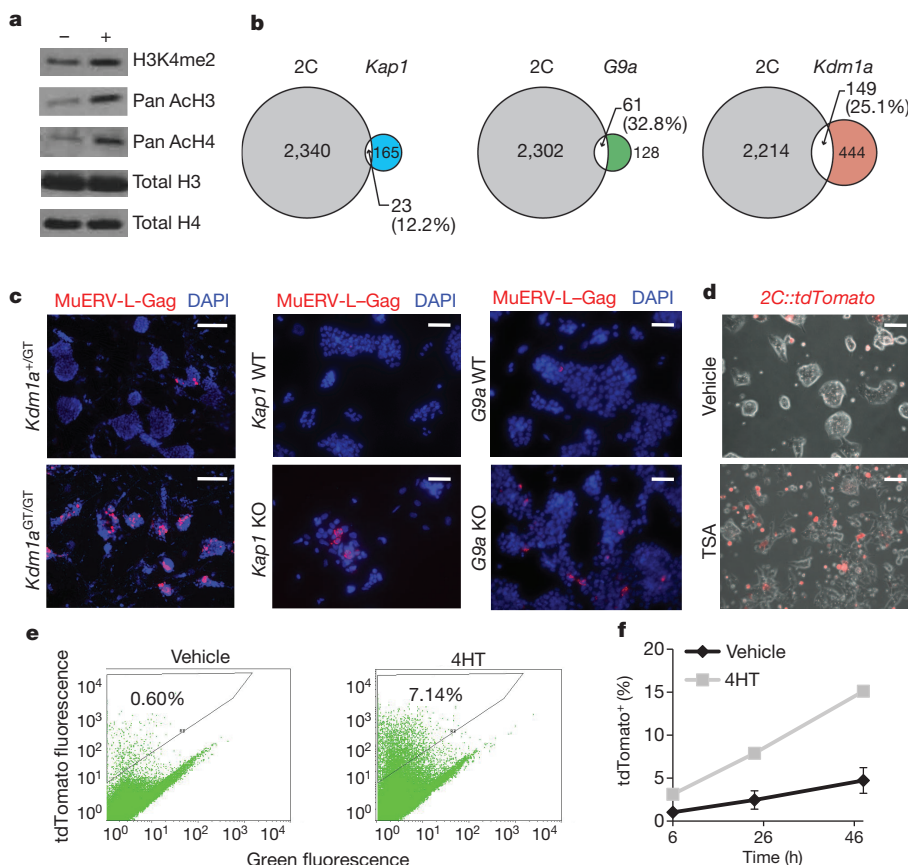


Figure 3 | The 2C state is associated with an active epigenetic signature and is antagonized by repressive chromatin-modifying enzymes.

a, 2C::tdTomato⁺ (+) and 2C::tdTomato⁻ (-) cells were collected by FACS and subjected to immunoblot analysis with indicated antibodies. H3K4me2, histone H3 dimethyl Lys 4; AcH3, acetylated histone H3. **b**, Pairwise comparisons of the number of genes activated in *Kap1*, *G9a* and *Kdm1a* mutant ES cells compared with genes activated in 2C embryos. **c**, ES cell lines homozygous for mutant alleles of *Kdm1a*, *Kap1* and *G9a*, and corresponding wild-type (WT) ES lines were immunostained with MuERV-L-Gag antibodies and counterstained with DAPI. GT, gene trap; KO, knockout. Scale bars, 50 μm. **d**, 2C::tdTomato ES cells were treated with 40 nM trichostatin A (TSA) for 24 h before imaging. Scale bars, 50 μm. **e**, *Kdm1a*^{Δ/Δ}; *Cre-ERT* ES cells containing a stably integrated 2C::tdTomato transgene were treated with vehicle or 4HT and subject to FACS analysis to determine the percentage of tdTomato⁺ cells. **f**, 2C::tdTomato; *Kdm1a*^{Δ/Δ}; *Cre-ERT* ES cells were treated with 4HT or vehicle for 24 h, then passaged for 72 h before collecting tdTomato⁺ cells by FACS. The percentage of tdTomato⁺ cells was plotted after increasing hours in culture. Error bars represent s.d., *n* = 3.

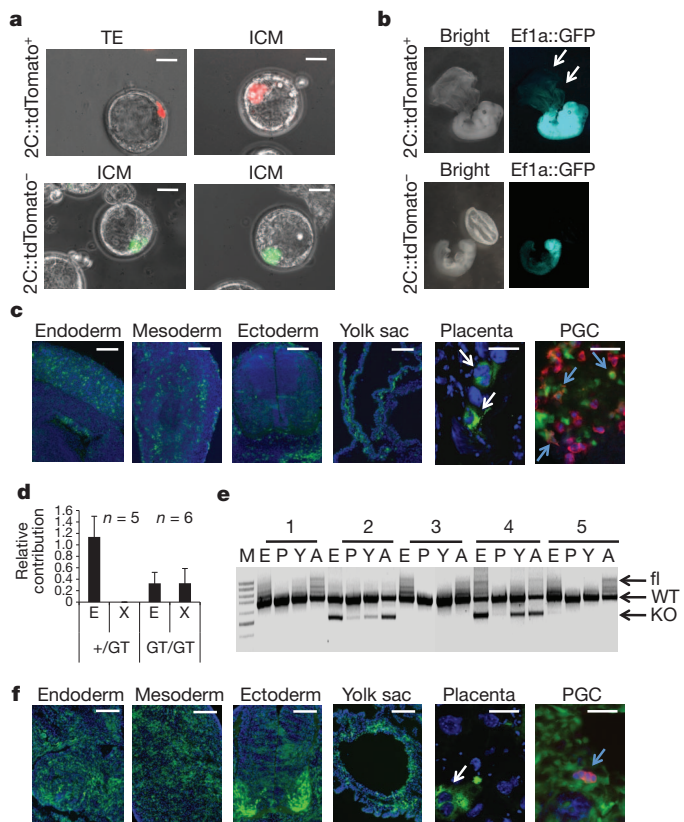


Figure 4 | Activation of the 2C state is associated with expanded potency in chimaeric embryos towards extraembryonic lineages. **a**, 2C::tdTomato⁺ or 2C::tdTomato⁻ cytomegalovirus (CMV)-GFP ES cells were injected into morula-stage embryos, which were then grown *in vitro*. The resulting blastocysts were imaged to visualize the position of injected cells in either the trophectoderm (TE) or ICM. Scale bars, 20 μ m. **b**, 2C::tdTomato⁺ or 2C::tdTomato⁻, Ef1a::GFP⁺ cells were injected into blastocysts that were then implanted into pseudopregnant females to generate chimaeric embryos. Arrows indicate 2C::tdTomato⁺, GFP⁺ cells contributing to the yolk sac and placenta. Bright denotes bright-field microscopy. **c**, 2C::tdTomato⁺, Ef1a::GFP⁺ cells contribute to embryonic endoderm, mesoderm, ectoderm, yolk sac, placental tissues (including giant trophoblast cells, white arrows) and primordial germ cells (PGCs, colabelled with anti-Ddx4 antibody in red, blue arrows). Scale bars, 500 μ m (endoderm, mesoderm, ectoderm and yolk sac) and 50 μ m (placenta and PGCs). **d**, Heterozygous (+/GT) or homozygous (GT/GT) *Kdm1a*- β -geo gene-trap ES cells were injected into wild-type blastocysts that were implanted into pseudopregnant females. (β -geo is a fusion of β -galactosidase and neomycin-resistance genes.) Embryonic (E) and extraembryonic (X) tissues were separated from chimaeric embryos, and subject to semiquantitative PCR with β -geo primers to determine the relative contribution of the injected cells to these lineages. Error bars represent s.e.m. **e**, A 1:1 mixture of *Kdm1a*^{fl/fl} control and *Kdm1a* knockout (KO) ES cells were co-injected into wild-type blastocysts. At embryonic day 12.5, chimaeric embryonic (E) tissue was separated from placenta (P), yolk sac (Y) and amnion (A) and subject to PCR to detect the floxed (fl) and knockout alleles of the injected cells relative to wild-type alleles of the resident injected embryo. M denotes PCR marker. **f**, *Kdm1a*^{GT/GT}, Ef1a::GFP⁺ cells contribute to embryonic endoderm, mesoderm, ectoderm, yolk sac, placental tissues (including giant trophoblast cells, white arrow) and primordial germ cells (PGCs, colabelled with anti-Ddx4 antibody in red, blue arrow). Scale bars, 500 μ m (endoderm, mesoderm, ectoderm and yolk sac) and 50 μ m (placenta and PGCs).

tdTomato⁻ cells that were pre-infected with a lentivirus encoding green fluorescent protein (GFP) from a constitutively active Ef1a promoter (Ef1a::GFP). tdTomato⁻ GFP⁺ cells contributed exclusively to embryonic tissues, whereas tdTomato⁺ GFP⁺ cells contributed to embryonic endoderm, ectoderm, mesoderm, the germ lineage as well as the yolk sac and placenta (Fig. 4b, c and Supplementary Fig. 6a, b). The extraembryonic contribution of the tdTomato⁺ GFP⁺ cells

included giant trophoblast cells of the placenta (Fig. 4c). Thus, the developmental potential of 2C::tdTomato⁺ cells includes embryonic plus extraembryonic tissues in contrast to most ES cells in culture, which are restricted to generating only embryonic cell types.

We next examined whether *Kdm1a* mutant ES lines, which contain higher frequencies of 2C::tdTomato⁺ cells, also had increased potency in mouse chimaera assays. As expected, *Kdm1a* heterozygous ES cells contributed exclusively to embryonic tissues (in five out of five chimaeric embryos) but never to extraembryonic tissues (Fig. 4d). By contrast, *Kdm1a* homozygous mutant ES cells generated both embryonic (in four out of six chimaeric embryos) and extraembryonic (in five out of six chimaeric embryos) tissues (Fig. 4d). To confirm the increased potential of *Kdm1a* mutant ES cells, we used a competition chimaera assay. We co-injected a 1:1 mixture of control *loxP*-flanked (floxed) *Kdm1a*^{fl/fl} and homozygous *Kdm1a* knockout ES cells into five wild-type blastocysts. PCR was then used to detect the appearance of *Kdm1a*^{fl/fl} or knockout cells in dissected tissues. We detected *Kdm1a*^{fl/fl} ES cells in the embryonic tissues and amnion, but not the yolk sac or placenta (Fig. 4e). By contrast, *Kdm1a* mutant ES cells contributed to embryonic tissues, the amnion, yolk sac and placental tissues, including giant trophoblast cells and primordial germ cells (Fig. 4e, f). Thus, the artificial activation of 2C genes achieved by removing *Kdm1a* is associated with expanded fate potential.

We have shown that 2C::tdTomato⁺ cells within ES cultures have increased potency, but it is unclear whether entrance into this state is essential for their long-term pluripotency. To test this possibility, we performed serial depletion of 2C-like ES cells by genetic ablation with diphtheria toxin (DTA). We generated ES lines by crossing 2C::ERT2-Cre-ERT2 mice (Supplementary Fig. 3a) with a Cre-responsive DTA allele (*ROSA::LSL-DTA*) and treated the ES line with 4HT for 20 passages (Supplementary Fig. 7a). We found that these 2C-depleted ES cultures were still capable of generating high contribution chimaeras (Supplementary Fig. 7b), although their differentiation was biased towards mesoderm and ectoderm lineages *in vitro* (Supplementary Fig. 7c). These data suggest that occasional entry into the 2C-like state might help to preserve the broad embryonic fate potential of ES cells.

Discussion

In mammalian development, the zygote and its daughter cells progress from totipotent cells capable of generating an entire mouse to more lineage-restricted inner and outer cells of the morula capable of generating embryonic or extraembryonic lineages, respectively. A key transcriptional feature of the totipotent cells is the onset of zygote genome activation in which the embryo switches from a maternal to a zygotic transcriptome. To mark cells at this early stage of embryonic development, we generated a reporter with the regulatory elements from the endogenous retrovirus MuERV-L, which is highly restricted to the zygote/2C stage. Surprisingly, we found that rare ES and iPS cells expressed the reporter. When we characterized these cells, we found that they lacked expression of the pluripotency proteins Oct4, Sox2 and Nanog. Instead, these rare cells expressed a large number of genes restricted to the 2C stage, and most importantly, were capable of forming both embryonic and extraembryonic lineages (Fig. 5a, b). Our studies identify a rare 2C-like cell in ES cultures that has expanded fate potential.

Although it is unclear how MuERV-L and other 2C genes regulate potency, several lines of evidence indicate that 2C-like cells are required for the health and maintenance of ES cultures. First, we found that nearly all cells enter into the 2C-like state over increasing passage. Second, when we depleted 2C-like ES cells from cell cultures we found that their differentiation characteristics were altered to generate more ectoderm and mesoderm derivatives. Third, functional studies of the *Zscan4* gene, found adjacent to a full-length MuERV-L element and highly enriched within 2C::tdTomato⁺ cells, have shown that it is required for the maintenance of telomeres within ES cultures¹⁴. Another important question that remains is whether the selection of

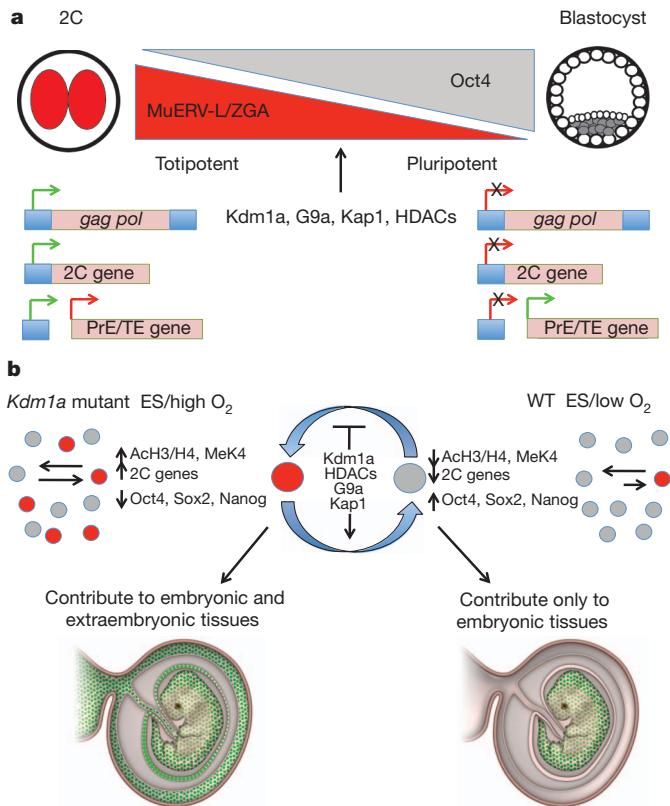


Figure 5 | Model of the role of the MuERV-L-LTR-linked 2C gene network in regulating embryonic potency. **a**, During zygote genome activation, a network of genes that use MuERV-L-LTRs as promoters is activated. This stage correlates with a period in which blastomeres are totipotent. As development progresses, the MuERV-L-LTR-linked 2C gene network is silenced by chromatin repressors, as the ICM segregates from the trophectoderm and primitive endoderm (PrE). HDACs, histone deacetylases. **b**, During the derivation of ES cells from blastocysts, a rare transient population of cells marked by the 2C::tdTomato reporter expresses high levels of 2C genes and low levels of pluripotency markers. In mouse chimaera assays, these cells contribute to embryonic and extraembryonic tissues (shown in green). Increasing the oxidative tension of ES cell cultures or deletion/inhibition of repressive histone-modifying enzymes alters the equilibrium between the 2C and ES states.

these special ES cells can be used for practical purposes, such as reprogramming somatic nuclei. This idea is supported by the finding that 2C genes are not properly activated in cloned embryos, and that reprogramming efficiency is enhanced by inhibition of histone deacetylases and Kdm1a, which repress the 2C state^{32–34}. Thus, overexpression of one or several MuERV-L-linked 2C genes or inhibition of other 2C gene repressors may be useful strategies to facilitate reprogramming. This possibility is supported by the recent finding that forced Zscan4 expression in fibroblasts enhances their iPS cell reprogramming efficiency³⁵.

Transposable elements are a major driving force of evolution. Our findings support the notion that the co-option of retrotransposable elements by cellular genes can act as an evolutionary mechanism for coordinately linking the expression of many genes^{15,29,36}. Transposon sequences have recently been shown to have a crucial role in rewiring gene regulatory networks in ES cells and in the endometrium that contributed to the evolution of pregnancy in mammals^{37,38}. It has also been speculated that ERVs were involved in the evolution of the placenta by providing fusogenic envelope genes adapted for formation of the syncytiotrophoblasts³⁹. We suggest that endogenous retroviruses, which are found in all placental mammals⁴⁰, may have had an equally important gene regulatory role in early mammalian development, by contributing to the specification of cell types and leading to the formation of placental tissues.

METHODS SUMMARY

2C::tdTomato was created by digesting the MuERV-L-LTR-Gag clone 9 (ref. 29) with MluI and HindIII, resulting in MuERV-L-LTR 1–730, and was ligated into pcDNA3 hygroTomato with the cytomegalovirus (CMV) promoter removed. To generate 2C::tdTomato ES cells, *Kdm1a*^{fl/fl}; *Cre-ERT* ES cells were transfected with 2C::tdTomato using Lipofectamine 2000 (Invitrogen) and selected with 150 µg ml⁻¹ hygromycin for 7 days. Colonies containing tdTomato⁺ cells were then picked and expanded. 2C::ERT2-*Cre-ERT2* was generated by replacing tdTomato with an ERT2-*Cre-ERT2* insert using EcoRI and NotI sites. DNA was linearized with MluI and AvrII sites before injection into embryos to generate transgenic mice. The resulting mice were mated with *ROSA::LSL-tdTomato* mice (JAX 007905), *ROSA::LSL-DTA* mice (JAX 010527) or *ROSA::LSL-LacZ* mice (gift from D. Anderson laboratory), and ES lines were derived using standard procedures. *Kdm1a*^{GT/GT}, *Kdm1a*^{fl/fl}, *Kap1* and *G9a* mutant ES cells were described previously^{29–31}. RNA-seq from oocytes and 2C embryos was performed by lysing litters of embryos (5–10 embryos) in prelude direct lysis buffer (Nugen), and amplifying RNA using the ovation RNA-seq system (Nugen) before library construction using the Tru-seq RNA sample prep kit (Illumina). Microarray, quantitative PCR with reverse transcription (qRT-PCR), immunostaining and chimaeric mouse injections were performed as described²⁹. All animal experiments we performed in accordance with the Salk Institute Institutional Animal Care and Use Committee guidelines.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 9 November 2011; accepted 21 May 2012.

Published online 13 June 2012.

- Tarkowski, A. K. Experiments on the development of isolated blastomeres of mouse eggs. *Nature* **184**, 1286–1287 (1959).
- Papaioannou, V. E., Mkandawire, J. & Biggers, J. D. Development and phenotypic variability of genetically identical half mouse embryos. *Development* **106**, 817–827 (1989).
- Cockburn, K. & Rossant, J. Making the blastocyst: lessons from the mouse. *J. Clin. Invest.* **120**, 995–1003 (2010).
- Latham, K. E. & Schultz, R. M. Embryonic genome activation. *Front. Biosci.* **6**, d748–d759 (2001).
- Schultz, R. M. The molecular foundations of the maternal to zygotic transition in the preimplantation embryo. *Hum. Reprod. Update* **8**, 323–331 (2002).
- Kan'ka, J. Gene expression and chromatin structure in the pre-implantation embryo. *Theriogenology* **59**, 3–19 (2003).
- Evans, M. J. & Kaufman, M. H. Establishment in culture of pluripotent cells from mouse embryos. *Nature* **292**, 154–156 (1981).
- Martin, G. R. Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proc. Natl Acad. Sci. USA* **78**, 7634–7638 (1981).
- Beddington, R. S. & Robertson, E. J. An assessment of the developmental potential of embryonic stem cells in the midgestation mouse embryo. *Development* **105**, 733–737 (1989).
- Niakan, K. K. *et al.* Sox17 promotes differentiation in mouse embryonic stem cells by directly regulating extraembryonic gene expression and indirectly antagonizing self-renewal. *Genes Dev.* **24**, 312–326 (2010).
- Hayashi, K., Lopes, S. M., Tang, F. & Surani, M. A. Dynamic equilibrium and heterogeneity of mouse pluripotent stem cells with distinct functional and epigenetic states. *Cell Stem Cell* **3**, 391–401 (2008).
- Singh, A. M., Hamazaki, T., Hankowski, K. E. & Terada, N. A heterogeneous expression pattern for Nanog in embryonic stem cells. *Stem Cells* **25**, 2534–2542 (2007).
- Chambers, I. *et al.* Nanog safeguards pluripotency and mediates germline development. *Nature* **450**, 1230–1234 (2007).
- Zalzman, M. *et al.* Zscan4 regulates telomere elongation and genomic stability in ES cells. *Nature* **464**, 858–863 (2010).
- Peaston, A. E. *et al.* Retrotransposons regulate host genes in mouse oocytes and preimplantation embryos. *Dev. Cell* **7**, 597–606 (2004).
- Evsikov, A. V. *et al.* Systems biology of the 2-cell mouse embryo. *Cytogenet. Genome Res.* **105**, 240–250 (2004).
- Kigami, D., Minami, N., Takayama, H. & Imai, H. MuERV-L is one of the earliest transcribed genes in mouse one-cell embryos. *Biol. Reprod.* **68**, 651–654 (2003).
- Svoboda, P. *et al.* RNAi and expression of retrotransposons MuERV-L and IAP in preimplantation mouse embryos. *Dev. Biol.* **269**, 276–285 (2004).
- Ribet, D. *et al.* Murine endogenous retrovirus MuERV-L is the progenitor of the “orphan” epsilon viruslike particles of the early mouse embryo. *J. Virol.* **82**, 1622–1625 (2008).
- Soudais, C. *et al.* Targeted mutagenesis of the transcription factor GATA-4 gene in mouse embryonic stem cells disrupts visceral endoderm differentiation *in vitro*. *Development* **121**, 3877–3888 (1995).
- Yagi, R. *et al.* Transcription factor TEAD4 specifies the trophectoderm lineage at the beginning of mammalian development. *Development* **134**, 3827–3836 (2007).
- Nishioka, N. *et al.* Tead4 is required for specification of trophectoderm in pre-implantation mouse embryos. *Mech. Dev.* **125**, 270–283 (2008).

23. Choo, K. B., Chen, H. H., Cheng, W. T., Chang, H. S. & Wang, M. *In silico* mining of EST databases for novel pre-implantation embryo-specific zinc finger protein genes. *Mol. Reprod. Dev.* **59**, 249–255 (2001).
24. Huang, C. J., Chen, C. Y., Chen, H. H., Tsai, S. F. & Choo, K. B. TDPOZ, a family of bipartite animal and plant proteins that contain the TRAF (TD) and POZ/BTB domains. *Gene* **324**, 117–127 (2004).
25. Zhang, W. *et al.* Zfp206 regulates ES cell gene expression and differentiation. *Nucleic Acids Res.* **34**, 4780–4790 (2006).
26. Ying, Q. L. *et al.* The ground state of embryonic stem cell self-renewal. *Nature* **453**, 519–523 (2008).
27. Ma, J., Svoboda, P., Schultz, R. M. & Stein, P. Regulation of zygotic gene activation in the preimplantation mouse embryo: global activation and repression of gene expression. *Biol. Reprod.* **64**, 1713–1721 (2001).
28. Wiekowski, M., Miranda, M., Nothias, J. Y. & DePamphilis, M. L. Changes in histone synthesis and modification at the beginning of mouse development correlate with the establishment of chromatin mediated repression of transcription. *J. Cell Sci.* **110**, 1147–1158 (1997).
29. Macfarlan, T. S. *et al.* Endogenous retroviruses and neighboring genes are coordinately repressed by LSD1/KDM1A. *Genes Dev.* **25**, 594–607 (2011).
30. Rowe, H. M. *et al.* KAP1 controls endogenous retroviruses in embryonic stem cells. *Nature* **463**, 237–240 (2010).
31. Yokochi, T. *et al.* G9a selectively represses a class of late-replicating genes at the nuclear periphery. *Proc. Natl Acad. Sci. USA* **106**, 19363–19368 (2009).
32. Suzuki, T., Minami, N., Kono, T. & Imai, H. Zygotically activated genes are suppressed in mouse nuclear transferred embryos. *Cloning Stem Cells* **8**, 295–304 (2006).
33. Shao, G. B. *et al.* Effect of trychostatin A treatment on gene expression in cloned mouse embryos. *Theriogenology* **71**, 1245–1252 (2009).
34. Li, W. *et al.* Generation of human-induced pluripotent stem cells in the absence of exogenous Sox2. *Stem Cells* **27**, 2992–3000 (2009).
35. Hirata, T. *et al.* Zscan4 transiently reactivates early embryonic genes during the generation of induced pluripotent stem cells. *Sci. Rep.* **2**, 208 (2012).
36. Feschotte, C. Transposable elements and the evolution of regulatory networks. *Nature Rev. Genet.* **9**, 397–405 (2008).
37. Kunarso, G. *et al.* Transposable elements have rewired the core regulatory network of human embryonic stem cells. *Nature Genet.* **42**, 631–634 (2010).
38. Lynch, V. J., Leclerc, R. D., May, G. & Wagner, G. P. Transposon-mediated rewiring of gene regulatory networks contributed to the evolution of pregnancy in mammals. *Nature Genet.* **43**, 1154–1159 (2011).
39. Dupressoir, A. *et al.* Syncytin-A knockout mice demonstrate the critical role in placentation of a fusogenic, endogenous retrovirus-derived, envelope gene. *Proc. Natl Acad. Sci. USA* **106**, 12127–12132 (2009).
40. B  nit, L., Lallemand, J. B., Casella, J. F., Philippe, H. & Heidmann, T. ERV-L elements: a family of endogenous retrovirus-like elements active throughout the evolution of mammals. *J. Virol.* **73**, 3301–3308 (1999).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We would like to thank members of the Pfaff laboratory for discussion; S. Andrews, D. Chambers, Y. Dayn, T.-C. Sung, J. Fitzpatrick, M. Joens, Y. Sigal, D. Gibbs and L. Ouyang for technical assistance, Y. Shinkai and D. Gilbert for the G9a mutant ES cells, and T. Heidmann for MuERV-L-Gag antibodies. This research was supported by the National Institute of Neurological Disorders and Stroke (R37NS037116) and the Marshall Heritage Foundation. T.S.M. and W.D.G. were supported by the California Institute for Regenerative Medicine and S.L.P. is an investigator of the Howard Hughes Medical Institute and Benjamin H. Lewis Chair in Neurobiology.

Author Contributions T.S.M. designed and performed all experiments with assistance from W.D.G., S.D., D.B. and K.L. under the supervision of S.L.P. D.T. generated *Kap1* mutant ES cells and H.M.R. and D.T. provided mRNA-seq data from these cells. A.F. and O.S. generated and provided iPS cell lines and lentivirus constructs. T.S.M., W.D.G. and S.L.P. wrote the manuscript.

Author Information Microarray and RNA-seq files have been submitted to the NCBI Gene Expression Omnibus database under accession GSE33923. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to S.L.P. (pfaff@salk.edu).

METHODS

RNA-seq. For RNA-seq analysis of early stage embryos, three independent litters of superovulated oocytes or naturally fertilized superovulated oocytes were collected and lysed directly in 2 µl of prelude direct lysis buffer (Nugen). RNA was then subject to amplification using the ovation RNA-seq system (Nugen). Amplified complementary DNA was fragmented using Covaris, and single-end (oocytes) or paired-end (2C embryos) libraries were then constructed using the mRNA-seq sample prep kit (Illumina) or Tru-seq RNA library construction kit (Illumina) starting with end repair. Sequencing was performed on either a Illumina genome analyzer (oocytes) or a Hi-Seq (2C embryos). A total of 72-base pair (bp) single-end reads (oocytes) or 100-bp paired-end reads (2C embryos) were aligned to the mouse genome using Bowtie, allowing up to three mismatches per alignment and up to 20 alignments per read, filtering out any read aligning in more than 20 locations. To compare the oocyte data with the 2C data, read lengths were cut down to 72 bp (from the 3' end). The oocyte data had an average of 33 million alignments per sample, whereas the 2C data had an average of 49 million alignments per sample. Read counts were quantified using a custom gene reference based on the University of California, Santa Cruz (UCSC) database. At each gene locus, all isoforms belonging to a single gene were fused into one transcript containing all exons from each isoform. Counts aligning in several locations were counted as a fraction of their total number of alignments at each location. Differential expression testing was performed with DESeq⁴¹. Genes with adjusted *P* values less than 0.05 were marked as significant. Chimaeric transcripts were identified using the spliced alignment data produced by Tophat. Tophat identifies exons based on alignment pileup, and it follows by aligning previously unaligned reads across potential splice junctions. We split the junction information into two lists, the left and right side of each junction, and compared with both the UCSC knowngene database for mm9 and the RepeatMasker database, also from the UCSC database. Only junctions that hit an exon of a known model on one end and a repeat element on the other were retained. GO analysis was performed using the David Bioinformatic Resource (<http://david.abcc.ncifcrf.gov/>)⁴².

For RNA-seq of 2C::tdTomato⁺ and 2C::tdTomato⁻ cells, and *Kdm1a*, *Kap1* and *G9a* knockout ES cells, sample libraries were prepared from 500 ng–5 µg of total RNA using the mRNA-seq sample prep kit (Illumina) or Tru-seq RNA library construction kit (Illumina) and then sequenced using consecutive 36 cycle sequencing kit on the genome analyzer (Illumina) or 100-bp paired-end reads on the Hi-Seq (Illumina). Raw sequence data was then aligned to the mouse genome using the short read aligner Bowtie and the default setting (two mismatches per 25 bp and up to 40 genomic alignments) (<http://bowtie-bio.sourceforge.net/index.shtml>). Reads per kilobase of exon model per million mapped reads (RPKM) values were also determined by Bowtie. For repetitive sequences, we aligned sequencing reads to the Repbase database using Bowtie (<http://www.girinst.org/repbase/index.html>). Differential expression was determined using DESeq as described earlier. To compare gene expression in *G9a*, *Kdm1a* and *Kap1* mutant ES cells, we identified upregulated genes by combining previously identified upregulated genes^{29–31} and our own DESeq analysis.

ES culture and generation of 2C::tdTomato and 2C::ERT2-Cre-ERT2 ES lines. The derivation and culture of *Kdm1a*^{GT/GT} and *Kdm1a*^{fl/fl}; *Cre-ERT* ES cells were described previously²⁹. The 2C::tdTomato construct was created by digesting the MuERV-L-LTR-Gag clone 9 in pGL3 basic with MluI and HindIII, resulting in MuERV-L-LTR 1-730, and was ligated into pcDNA3 hygro tdTomato digested with MluI and HindIII (to remove the CMV promoter). To generate 2C::tdTomato ES cells, *Kdm1a*^{fl/fl}; *Cre-ERT* ES cells were transfected with 2C::tdTomato using Lipofectamine 2000 (Invitrogen) and selected with 150 µg ml⁻¹ hygromycin for 7 days. Colonies containing tdTomato⁺ cells were then picked and expanded. 2C::tdTomato ES cells were also derived from a transgenic mouse generated by pronuclear injection of the 2C::tdTomato ES line. 2C::ERT2-Cre-ERT2 was generated by replacing tdTomato with an ERT2-Cre-ERT2 insert using EcoRI and NotI sites. DNA was linearized with MluI and AvrII sites before injection into embryos to generate transgenic mice. The resulting mice were mated with ROSA::LSL-tdTomato mice (JAX 007905), ROSA::LSL-DTA mice (JAX 010527) or ROSA::LSL-LacZ mice (gift from D. Anderson laboratory), and ES cell lines were derived using standard procedures.

Kap1 and *G9a* mutant ES cells were described previously^{30,31}. To recombine and delete the *Kdm1a*, *Kap1* and *G9a* floxed alleles, cells were treated with 1 µM 4HT for 24 h. Cells were collected at a minimum of 48 h later to allow for loss of residual protein. To activate the 2C::ERT2-Cre-ERT2 transgene, cells were maintained in 1 µM 4HT and fed daily. tdTomato⁺ and tdTomato⁻ cells were counted using a FACScan and sorted using a FACSDiVA. For differentiation assays, ES cells were grown in suspension in the absence of leukaemia inhibitory factor as described²⁹.

Immunofluorescence staining and microscopy. ES and iPS cells were plated on gelatinized glass coverslips on primary mouse embryonic fibroblasts. Cells were fixed with 4% paraformaldehyde (PFA) for 10 min, followed by washing with PBS-T (0.05% tween). Cells were blocked in PBS-T containing 3% BSA for 10 min and stained with primary antibody for 1 h at room temperature. Antibodies used: mouse anti-Kap1 (Abcam, 1:1,000); mouse anti-Oct3/4 (Santa Cruz, 1:500); rabbit anti-MuERV-L-Gag (gift from T. Heidmann laboratory, 1:2,000); rat anti-E-cadherin (Abcam, 1:500); rabbit anti-Pan-acetylated histone H3 (Upstate, 1:1,000); rabbit anti-Pan-acetyl H4 (Upstate, 1:1,000); and rabbit anti-H3K4me2 (Abcam, 1:1,000). After washing three times for 10 min with PBS-T, cells were stained with secondary antibody (1:1,000 anti-mouse, -rat or -rabbit IgG Alexa fluor 488, 55 or 647) for 1 h at room temperature and washed again three times with PBS-T. Coverslips were stained with DAPI in PBS for 5 min before inverting onto slides in mounting medium. Cells were then imaged using either an Olympus FV1000 confocal microscope and a ×60 oil objective, or a Zeiss Axioskop 2 epifluorescence microscope and ×40 objective. Quantification of histone stains was performed with Fluoview. Preimplantation embryos were stained as described with minor modifications. Embryos were fixed in 4% PFA for 30 min and permeabilized in 0.25% Triton for 20 min, before blocking in 10% FBS for 1 h in 0.1% Triton-PBS. Primary antibodies were incubated overnight at 4 °C in blocking buffer. Subsequent washes and secondary antibody incubations were at room temperature in 0.1% Triton-PBS.

In situ hybridization. A MuERV-L probe was generated by PCR from mouse ES cDNA using the forward primer 5'-CCATCCCTGTCATTGCTCA-3' and reverse primer 5'-CCTTTTCCACCCCTTGATT-3', and cloned into the PCR2.1 TOPO vector. A digoxigenin (DIG)-labelled probe was prepared using *in vitro* transcription with the T7 polymerase. ES cell samples were fixed in 4% PFA, digested for 2 min with proteinase K, washed with PBS, acetylated and hybridized with denatured probe overnight at 68 °C. After washing with 5× saline sodium citrate (SSC) and 0.2× SSC, DIG-labelled probe was visualized using an anti-DIG antibody coupled to alkaline phosphatase.

Immunoblotting. Whole cell extracts were prepared by pelleting ES cells at 200g and resuspending in 1:5 volume of 1% NP40 lysis buffer containing 10 mM Tris, 150 mM NaCl and 1× protease inhibitors. To solubilize histones, extracts were also sonicated using a bioruptor on the high setting for 10 min. Ten to fifty micrograms of total protein in LDS sample buffer (Invitrogen) was then loaded onto a 4–12% NuPage gel (Invitrogen), electrophoresed at 200 V for 60 min, and transferred to nitrocellulose membranes at 30 V for 90 min. Membranes were blocked in PBS-T containing 5% non-fat dry milk. Primary antibodies were incubated overnight at 4 °C. Antibodies used: rabbit anti-GAPDH (Santa Cruz, 1:1,000); rabbit anti-MuERV-L-Gag (gift from T. Heidmann laboratory, 1:1,000); anti-Pan-AcH3 (Upstate, 1:1,000); anti-Pan-AcH4 (Upstate, 1:1,000); anti-H3K4me2 (Abcam, 1:500); anti-H4 (Novus, 1:1,000); and anti-H3 (Novus, 1:500). After washing extensively with PBS-T, secondary antibodies (anti-rabbit or -mouse horseradish peroxidase (HRP) conjugate, 1:10,000 dilution) were incubated for 1 h at room temperature. After washing extensively with PBS-T and water, blots were developed using ECL plus detection system (Amersham).

Bisulphite sequencing. ES cells were lysed in lysis buffer (0.1 M Tris, pH 8.5, 5 mM EDTA, 0.2% SDS and 0.2 M NaCl) containing proteinase K (Roche) for 1 h at 55 °C, followed by treatment with DNase-free RNase for 30 min at 37 °C. DNA was then sonicated briefly and purified using Qiagen PCR purification columns. Bisulphite conversion of genomic DNA was carried out using the epitect bisulphite kit (Qiagen). Bisulphite-converted DNA was then PCR-amplified using accuprime Taq polymerase (Invitrogen) followed by TOPO TA cloning (Invitrogen). At least 10 individual clones per primer pair were sequenced (Eton Bio). Primer sequences were described previously²⁹.

qRT-PCR. For quantitative PCR with reverse transcription (qRT-PCR) analysis, first-strand cDNA was generated from up to 5 µg total RNA using Superscript III (Invitrogen) and polydT or random hexamer priming. qPCR was performed using SYBR green master mix (Applied Biosystems) in 96-well dishes in triplicate and repeated with at least two biological replicates with similar results. Standard curves were generated for each primer pair (described previously²⁹) and expression levels were plotted relative to *Gapdh* (in arbitrary units).

Microarray. Total RNA was prepared from 2C::tdTomato⁺ and 2C::tdTomato⁻ cells using RNeasy kits (Qiagen). Labelling of 100 ng of total RNA was performed using the whole transcript sense target labelling assay kit (Affymetrix) before hybridization to genechip mouse gene 1.0 ST arrays. Probeset normalization and summarization were prepared using robust multichip analysis (RMA) in expression console (Affymetrix).

Mouse chimaera assay. ES cells were injected into either embryonic day (E) 2.5 or E3.5 C57Bl/6J embryos, and cultured *in vitro* or implanted into pseudopregnant

females. For PCR assays, dissected tissues were placed in lysis buffer (1% SDS, 150 mM NaCl, 10 mM Tris, pH 8.0, 1 mM EDTA, pH 8.0) containing proteinase K overnight at 55 °C. DNA was then isolated by phenol–chloroform extraction and ethanol precipitation, followed by PCR analysis with primers designed to amplify the β -geo cassette or the wild-type *Kdm1a* floxed allele. For embryo imaging, chimaeric mice were collected between E9.5 and E12.5 and fixed with 4% PFA for 2 h, washed extensively in PBS overnight, incubated in 30% sucrose

for 4 h, and frozen on dry ice in OCT. Cryosections were then taken and stained with DAPI before imaging.

41. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **11**, R106 (2010).
42. Huang da, W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols* **4**, 44–57 (2009).

The signature of the first stars in atomic hydrogen at redshift 20

Eli Visbal^{1,2}, Rennan Barkana³, Anastasia Fialkov³, Dmitriy Tselikhovich⁴ & Christopher M. Hirata⁵

Dark and baryonic matter moved at different velocities in the early Universe, which strongly suppressed star formation in some regions¹. This was estimated² to imprint a large-scale fluctuation signal of about two millikelvin in the 21-centimetre spectral line of atomic hydrogen associated with stars at a redshift of 20, although this estimate ignored the critical contribution of gas heating due to X-rays^{3,4} and major enhancements of the suppression. A large velocity difference reduces the abundance of haloes^{1,5,6} and requires the first stars to form in haloes of about a million solar masses^{7,8}, substantially greater than previously expected^{9,10}. Here we report a simulation of the distribution of the first stars at redshift 20 (cosmic age of around 180 million years), incorporating all these ingredients within a 400-megaparsec box. We find that the 21-centimetre hydrogen signature of these stars is an enhanced (ten millikelvin) fluctuation signal on the hundred-megaparsec scale, characterized² by a flat power spectrum with prominent baryon acoustic oscillations. The required sensitivity to see this signal is achievable with an integration time of a thousand hours with an instrument like the Murchison Wide-field Array¹¹ or the Low Frequency Array¹² but designed to operate in the range of 50–100 megahertz.

The relative velocity between the dark matter and baryons also reduces the gas content of each halo. Previous work² assumed that this reduces star formation, but it mainly affects smaller haloes that do not form stars^{13,14}. Another important consideration when selecting which era to focus on for observations of early stars is timing, because on the one hand, early times bring us closer to the primeval era of the very first stars^{9,10,14,15}, but on the other hand, the cosmological 21-cm signal is obscured by the foreground (mainly Galactic synchrotron), which is brighter at longer wavelengths (corresponding to higher redshifts). Unlike the fluctuations at redshift $z \approx 20$ from inhomogeneous gas heating, previously considered sources² produce smaller fluctuations and are likely to be effective only at $z \approx 30$ (ref. 16).

We use a hybrid method to produce realistic, three-dimensional images of the expected global distribution of the first stars. We use the known statistical properties of the initial perturbations of density and of the relative dark matter to baryon velocity to generate a realistic sample universe on large, linear scales. Then, we calculate the stellar content of each pixel using analytical models and the results of small-scale numerical simulations. In this approach we build upon previous hybrid methods used for high-redshift galaxy formation^{1,2,17}, and include a fit¹⁴ to recent simulation results on the effect of the relative velocity^{7,8} (for further details, see Supplementary Information section 1). We note that numerical simulations (even if limited to following gravity) cannot on their own cover the full range of scales needed to find the large-scale distribution of high-redshift galaxies¹⁸.

We assume standard initial perturbations (for example, from a period of inflation), where the density and velocity components are Gaussian random fields. Velocities are coherent on larger scales than the density, owing to the extra factor of $1/k$ in the velocity from the continuity

equation that relates the two fields (where k is the wavenumber). Indeed, velocity fluctuations have significant power over the range $k \approx 0.01$ – 0.5 per megaparsec, with prominent baryon acoustic oscillations¹.

We find a remarkable cosmic web (Fig. 1), reminiscent of that seen in the distribution of massive galaxies in the present Universe^{19–21}. The large coherence length of the velocity makes it the dominant factor (relative to density) in the large-scale pattern. The resulting enhanced structure on 100-megaparsec scales becomes especially notable at the highest redshifts (Fig. 2). This large-scale structure has momentous implications for cosmology at high redshift and for observational prospects. As the first stars formed, their radiation (plus emission from stellar remnants) produced feedback that radically affected both the intergalactic medium and the character of newly forming stars. Before reionization, three major transitions are expected, due to energetic photons. Lyman- α photons couple the hyperfine levels of hydrogen to the kinetic temperature and thus make possible 21-cm observations of this cosmic era, while X-rays heat the cosmic gas³. Meanwhile, Lyman-Werner (L-W) photons dissociate molecular hydrogen and eventually end the era of primordial star formation driven by molecular cooling²², leading to the dominance of larger haloes (which are more weakly affected by the relative velocities). Owing to the strong spatial fluctuations in the stellar sources¹⁸, these radiation backgrounds are inhomogeneous and should produce rich structure in 21-cm maps^{4,16,23,24}.

These radiation backgrounds have effective horizons of the order of a hundred megaparsecs, due to redshift, optical depth and time delay effects. Thus, the relative velocity effect on the stellar distribution leads to large-scale fluctuations in the radiation fields. This substantially alters the feedback environment of the first stars, making it far more inhomogeneous than previously thought. Observationally, these degree-scale fluctuations will affect various cosmic radiation backgrounds, and in particular the history of 21-cm emission and absorption (Fig. 3), which depends on the timing of the three radiative transitions. Although it is still significantly uncertain, the 21-cm coupling due to Lyman- α radiation is expected to occur rather early, with the X-ray heating fluctuations occurring later and probably overlapping with significant small-halo suppression due to L-W radiation (see Supplementary Information section 3 and Supplementary Fig. 1). Thus, we focus on the fluctuations due to X-ray heating at redshift 20, assuming that Lyman- α coupling has already saturated while bracketing the effect of the L-W flux by considering the two limiting cases where the L-W transition has either not yet begun or has already saturated.

Fluctuations on large scales are easier to observe, because 21-cm arrays rapidly lose sensitivity with increasing resolution²⁵. For fixed co-moving pixels, going from $z = 10$ to $z = 20$ increases the thermal noise per pixel by a factor of 30 (in the power spectrum), but this is more than compensated for if the required co-moving resolution is four times lower than at $z = 10$. In the case of negligible L-W flux, the

¹Jefferson Laboratory of Physics, Harvard University, Cambridge, Massachusetts 02138, USA. ²Institute for Theory and Computation, Harvard University, 60 Garden Street, Cambridge, Massachusetts 02138, USA. ³Raymond and Beverly Sackler School of Physics and Astronomy, Tel Aviv University, Tel Aviv 69978, Israel. ⁴California Institute of Technology, Mail Code 350-17, Pasadena, California 91125, USA. ⁵California Institute of Technology, Mail Code 249-17, Pasadena, California 91125, USA.

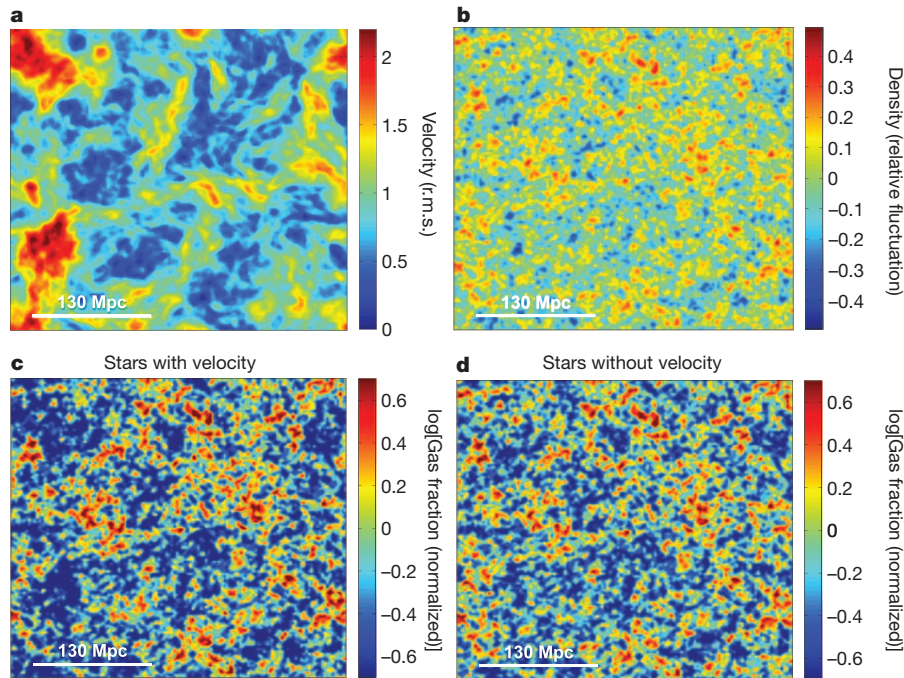


Figure 1 | The effect of relative velocity on the distribution of star-forming haloes at $z = 20$. A two-dimensional slice (thickness 3 megaparsecs) of a simulated cubic volume of 384 megaparsecs (co-moving) per side at $z = 20$. **b** and **d**, To illustrate previous expectations we show the overdensity (the relative fluctuation in density) (**b**) and the relative fluctuation of the gas fraction in star-forming haloes with the effect of density only (**d**). **a** and **c**, To illustrate the new predictions we show the magnitude of the relative baryon to dark-matter velocity (**a**), and the relative fluctuation of the gas fraction in star-forming haloes including the effect of relative velocity (**c**). The relative velocity

is given in units of the root-mean-square value. For the gas fraction (**c** and **d**), the colours correspond to the logarithm of the fraction normalized by the mean values, 0.0012 and 0.0021 for the case with and without the velocity effect, respectively; for ease of comparison, the scale in each plot ranges from 1/5 to 5 times the mean. In each panel, we indicate the scale of 130 co-moving megaparsecs, which corresponds to the large-scale peak in the 21-cm power spectrum (see Fig. 4). The no-velocity gas fraction map (**d**) is a biased version of the density map, whereas the velocity effect (**c**) increases the large-scale power and the map's contrast, producing larger, emptier voids.

relative velocity effect boosts the power spectrum on a scale of $2\pi/k = 130$ megaparsecs (0.66° at $z = 20$) by a factor of 3.8, leading to 11-millikelvin fluctuations on this scale and an overall flat power spectrum with a prominent signature of baryon acoustic oscillations (Fig. 4). If, however, the L–W transition has already saturated, the power spectrum is even higher (for example, 13 millikelvin on the above scale) owing to the dominance of larger haloes (characterized by efficient atomic cooling) that are more highly biased; in this case, the effect of the streaming velocities is suppressed, reducing the oscillatory signature and steepening the power spectrum. We thus predict a strong, observable signal from heating fluctuations, regardless of the precise timing of the L–W transition, with the signal's shape indicating the relative abundance of small versus large galaxies.

In general, the 21-cm fluctuation amplitude at a given redshift can be reduced by making galactic haloes less massive (and thus less strongly clustered) or by increasing the X-ray efficiency (thus heating the cosmic gas past the temperature range that affects the 21-cm emission). Thus, the characteristic shape that we predict is essential for resolving this degeneracy and allowing a determination of the properties of the early galaxies. Moreover, similar observations over the full $\Delta z \approx 6$ redshift range of significant heating fluctuations could actually detect the slow advance of the L–W feedback process, during which the power spectrum continuously changes shape, gradually steepening as the baryon acoustic oscillation signature weakens towards low redshift.

The exciting possibility of observing the 21-cm power spectrum from galaxies at $z \approx 20$ should stimulate observational efforts focused

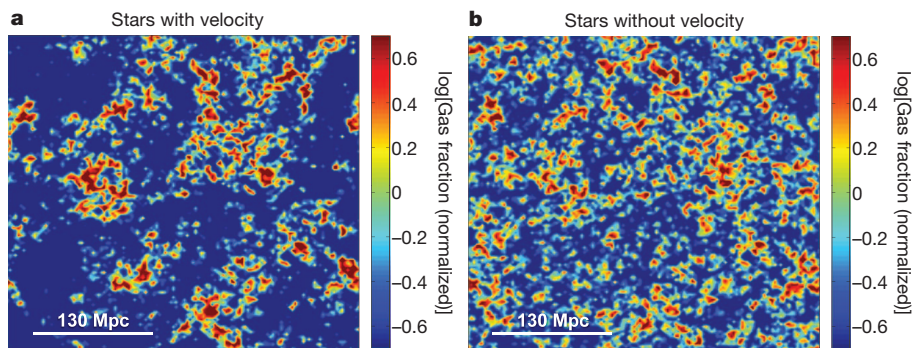


Figure 2 | The effect of relative velocity on the distribution of star-forming haloes at $z = 40$. The gas fraction in star-forming haloes in the same two-dimensional slice as in Fig. 1 but at $z = 40$, with (**a**) and without (**b**) the relative velocity effect. The colours correspond to the logarithm of the gas fraction normalized by the mean values, 1.5×10^{-8} and 1.1×10^{-7} for the case with

and without the relative velocity effect, respectively. The trends seen at $z = 20$ with the velocity effect are much stronger at $z = 40$, with hundred-megaparsec scales characterized by nearly isolated star-forming concentrations surrounded by deep voids; this implies a much more complex stellar feedback history, with the star-forming centres affected long before the voids.

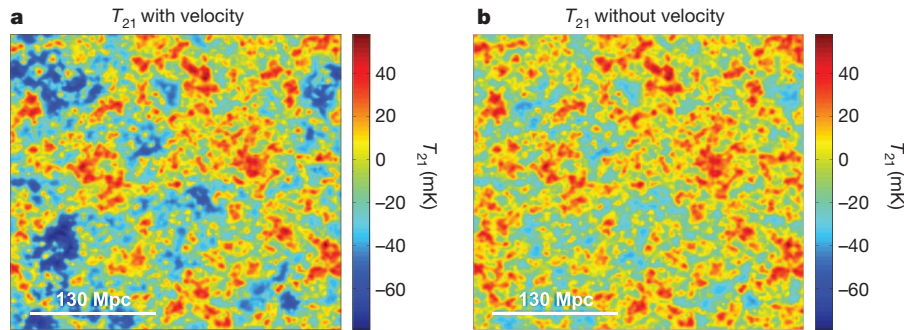


Figure 3 | The effect of relative velocity on the 21-cm brightness temperature at $z = 20$. The 21-cm brightness temperature (in units of millikelvin) in the same two-dimensional slice as in Fig. 1 at $z = 20$ with (a) and

on this early epoch. Such observations would push well past the current frontier of cosmic reionization ($z \approx 10$, $t \approx 480$ Myr) for galaxy searches²⁶ and 21-cm arrays²⁵. Detecting the remarkable velocity-caused baryon acoustic oscillation signature (which is much more prominent than its density-caused low-redshift counterpart²⁷) would confirm the major influence on galaxy formation of the initial velocity difference set at cosmic recombination. Measuring the abundance of

without (b) the relative velocity effect. For ease of comparison, both plots use a common scale that ranges from -79 millikelvin to $+58$ millikelvin; the no-velocity case is smoother and does not reach below -42 millikelvin.

million-solar-mass haloes would also probe primordial density fluctuations on approximately 20-kiloparsec scales, an order of magnitude below current constraints. This could lead to new limits on models with suppressed small-scale power such as warm dark matter²⁸.

Received 3 January; accepted 24 April 2012.

Published online 20 June 2012.

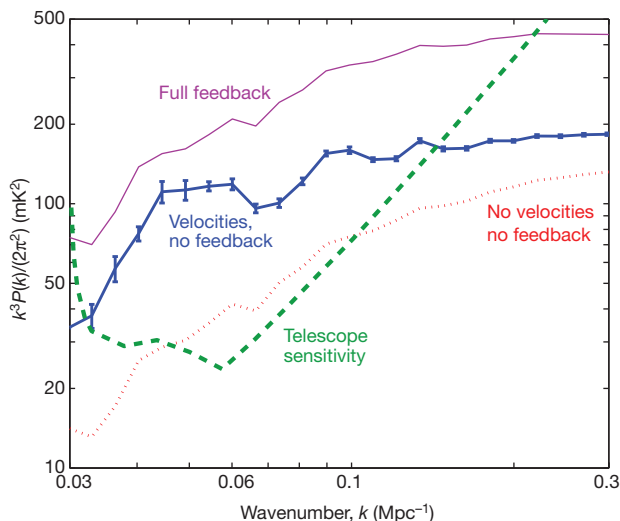


Figure 4 | The signature of relative velocity in the 21-cm power spectrum at $z = 20$. Power spectrum of the 21-cm brightness temperature fluctuations versus wavenumber, at the peak of the X-ray heating transition at $z = 20$. We show the prediction for the case of a late L–W transition for which the L–W feedback is still negligible at $z = 20$ (blue solid curve); this no-feedback case shows a strong velocity effect. This is well above the projected 1σ telescope sensitivity²⁹ (green dashed curve) based on 1,000-hour observations with an instrument like the Murchison Wide-field Array or the Low Frequency Array but designed to operate at 50–100 MHz, where we include an estimated degradation factor due to foreground removal³⁰ (see Supplementary Information section 4 for details). Future experiments like the Square Kilometer Array should reach a sensitivity that is better by an order of magnitude²⁹. We also show the prediction for an early L–W transition that has already saturated by $z = 20$, in which case the power spectrum is essentially unaffected by the velocities (purple solid curve). These two feedback cases bracket the possible range. We show for comparison the previous expectation for the no-feedback case, without the velocity effect (red dotted curve). The velocity effect makes it significantly easier to detect the signal, and also creates a clear signature by flattening the power spectrum and increasing the prominence of the baryon acoustic oscillations (which are more strongly imprinted in the velocity than in the density fluctuations). Each plotted result is the mean of 20 realizations of our full box (standard deviation error bars are shown in the main case). In this plot we have fixed the heating transition at $z = 20$ for easy comparison among the various cases. In cases with effective feedback, the heating transition as well as the main portion of the feedback itself will be delayed to a lower redshift, making the signal more easily observable.

1. Tseliakhovich, D. & Hirata, C. Relative velocity of dark matter and baryonic fluids and the formation of the first structures. *Phys. Rev. D* **82**, 083520 (2010).
2. Dalal, N., Pen, U.-L. & Seljak, U. Large-scale BAO signatures of the smallest galaxies. *J. Cosmol. Astroparticle Phys.* **11**, 007 (2010).
3. Madau, P., Meiksin, A. & Rees, M. J. 21 centimeter tomography of the intergalactic medium at high redshift. *Astrophys. J.* **475**, 429–444 (1997).
4. Pritchard, J. R. & Furlanetto, S. 21-cm fluctuations from inhomogeneous X-ray heating before reionization. *Mon. Not. R. Astron. Soc.* **376**, 1680–1694 (2007).
5. Maio, U., Koopmans, L. V. E. & Ciardi, B. The impact of primordial supersonic flows on early structure formation, reionization and the lowest-mass dwarf galaxies. *Mon. Not. R. Astron. Soc.* **412**, L40–L44 (2011).
6. Naoz, S., Yoshida, N. & Gnedin, N. Y. Simulations of early baryonic structure formation with stream velocity: I. Halo abundance. *Astrophys. J.* **747**, 128 (2012).
7. Stacy, A., Bromm, V. & Loeb, A. Effect of streaming motion of baryons relative to dark matter on the formation of the first stars. *Astrophys. J.* **730**, L1 (2011).
8. Greif, T., White, S., Klessen, R. & Springel, V. The delay of population III star formation by supersonic streaming velocities. *Astrophys. J.* **736**, 147 (2011).
9. Abel, T., Bryan, G. L. & Norman, M. L. The formation of the first star in the Universe. *Science* **295**, 93–98 (2002).
10. Bromm, V., Coppe, P. S. & Larson, R. B. Forming the first stars in the Universe: The fragmentation of primordial gas. *Astrophys. J.* **527**, L5–L8 (1999).
11. Bowman, J. D., Morales, M. F. & Hewitt, J. N. Foreground contamination in interferometric measurements of the redshifted 21 cm power spectrum. *Astrophys. J.* **695**, 183–199 (2009).
12. Harker, G. et al. Power spectrum extraction for redshifted 21-cm Epoch of Reionization experiments: the LOFAR case. *Mon. Not. R. Astron. Soc.* **405**, 2492–2504 (2010).
13. Tseliakhovich, D., Barkana, R. & Hirata, C. Suppression and spatial variation of early galaxies and minihalos. *Mon. Not. R. Astron. Soc.* **418**, 906–915 (2011).
14. Fialkov, A., Barkana, R., Tseliakhovich, D. & Hirata, C. Impact of the Relative Motion between Dark Matter and Baryons on the First Stars. *Mon. Not. R. Astron. Soc.* (submitted); preprint at <http://arxiv.org/abs/1110.2111>.
15. Naoz, S., Noter, S. & Barkana, R. The first stars in the Universe. *Mon. Not. R. Astron. Soc.* **373**, L98–L102 (2006).
16. Holzbauer, L. N. & Furlanetto, S. R. Fluctuations in the high-redshift Lyman-Werner and Lyman-alpha radiation backgrounds. *Mon. Not. R. Astron. Soc.* **419**, 718–731 (2012).
17. Mesinger, A., Furlanetto, S. & Cen, R. 21CMFAST: a fast, seminumerical simulation of the high-redshift 21-cm signal. *Mon. Not. R. Astron. Soc.* **411**, 955–972 (2011).
18. Barkana, R. & Loeb, A. Unusually large fluctuations in the statistics of galaxy formation at high redshift. *Astrophys. J.* **609**, 474–481 (2004).
19. Aihara, H. et al. The eighth data release of the Sloan Digital Sky Survey: first data from SDSS-III. *Astrophys. J. Suppl.* **193**, 29 (2011); erratum **195**, 26 (2011).
20. Colless, M. et al. The 2dF Galaxy Redshift Survey: spectra and redshifts. *Mon. Not. R. Astron. Soc.* **328**, 1039–1063 (2001).
21. Springel, V., Frenk, C. S. & White, S. D. M. The large-scale structure of the Universe. *Nature* **440**, 1137–1144 (2006).
22. Haiman, Z., Rees, M. J. & Loeb, A. Destruction of molecular hydrogen during cosmological reionization. *Astrophys. J.* **476**, 458–463 (1997); erratum **484**, 985 (1997).
23. Barkana, R. & Loeb, A. Detecting the earliest galaxies through two new sources of 21 centimeter fluctuations. *Astrophys. J.* **626**, 1–11 (2005).
24. Naoz, S. & Barkana, R. Detecting early galaxies through their 21-cm signature. *Mon. Not. R. Astron. Soc.* **385**, L63–L67 (2008).
25. Furlanetto, S. R., Oh, S. P. & Briggs, F. H. Cosmology at low frequencies: the 21 cm transition and the high-redshift Universe. *Phys. Rep.* **433**, 181–301 (2006).

26. Bouwens, R. J. *et al.* A candidate redshift $z \approx 10$ galaxy and rapid changes in that population at an age of 500 Myr. *Nature* **469**, 504–507 (2011).
27. Anderson, L. *et al.* The clustering of galaxies in the SDSS-III Baryon Oscillation Spectroscopic Survey: baryon acoustic oscillations in the Data Release 9 Spectroscopic Galaxy Sample. Preprint at <http://arxiv.org/abs/1203.6594> (2012).
28. Barkana, R., Haiman, Z. & Ostriker, J. P. Constraints on warm dark matter from cosmological reionization. *Astrophys. J.* **558**, 482–496 (2001).
29. McQuinn, M., Zahn, O., Zaldarriaga, M., Hernquist, L. & Furlanetto, S. R. Cosmological parameter estimation using 21 cm radiation from the epoch of reionization. *Astrophys. J.* **653**, 815–834 (2006).
30. Liu, A. & Tegmark, M. How well can we measure and understand foregrounds with 21-cm experiments? *Mon. Not. R. Astron. Soc.* **419**, 3491 (2012).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by the Israel Science Foundation (for R.B., and E.V.'s stay at Tel Aviv University) and by the European Research Council (for A.F.). D.T. and C.M.H. were supported by the US Department of Energy and the National Science Foundation. C.M.H. is also supported by the David & Lucile Packard Foundation.

Author Contributions R.B. initiated the project, and E.V. made the computations and figures by developing a code, parts of which were based on codes supplied by A.F., D.T. and C.M.H. The text was written by R.B. and edited by the other authors. A.F. added the L–W module for Supplementary section 3 and made Supplementary Fig. 1.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to E.V. (evibal@fas.harvard.edu) or R.B. (barkana@wise.tau.ac.il).

Rapid disappearance of a warm, dusty circumstellar disk

Carl Melis¹, B. Zuckerman², Joseph H. Rhee³, Inseok Song⁴, Simon J. Murphy⁵ & Michael S. Bessell⁵

Stars form with gaseous and dusty circumstellar envelopes, which rapidly settle into disks that eventually give rise to planetary systems. Understanding the process by which these disks evolve is paramount in developing an accurate theory of planet formation that can account for the variety of planetary systems discovered so far. The formation of Earth-like planets through collisional accumulation of rocky objects within a disk has mainly been explored in theoretical and computational work in which post-collision ejecta evolution typically is ignored^{1–3}, although recent work has considered the fate of such material⁴. Here we report observations of a young, Sun-like star (TYC 8241 2652 1) where infrared flux from post-collisional ejecta has decreased drastically, by a factor of about 30, over a period of less than two years. The star seems to have gone from hosting substantial quantities of dusty ejecta, in a region analogous to where the rocky planets orbit in the Solar System, to retaining at most a meagre amount of cooler dust. Such a phase of rapid ejecta evolution has not been previously predicted or observed, and no currently available physical model satisfactorily explains the observations.

TYC 8241 2652 1 (stellar parameters are reported in Table 1), was found as part of a survey to identify main-sequence stars with excess emission at mid- and far-infrared wavelengths. To accomplish this goal, we cross-correlated the Tycho-2 catalogue⁵ with those of the Infrared Astronomical Satellite, AKARI⁶ and the Wide-field Infrared Survey Explorer⁷ (WISE) and performed our own observations using the Thermal-Region Camera Spectrograph⁸ (T-ReCS) at the Gemini South telescope. Figure 1 and Table 2 show how the 11- μm excess emission of this source evolved from being a factor of ~ 30 times the stellar photosphere flux before 2009 to being ~ 13 times the photospheric flux in mid 2009 and being barely detectable in 2010 (details regarding each measurement can be found in Table 2). The pre-2009 measurements indicate significant mid-infrared excess emission and, hence, that warm dusty material orbited in the star's inner planetary system (Fig. 1 and Table 2). Remarkably, two epochs of WISE measurements show that the excess mid-infrared emission has all but disappeared, leaving only a weak (~ 3 times the stellar photosphere) excess at a wavelength of 22 μm (Fig. 1 and Table 2). We note that the two WISE epochs have a time separation of roughly six months and yet still report identical flux levels. Measurements made after the WISE epochs using the SpeX spectrograph at the NASA Infrared Telescope Facility^{9–11}, the Photodetector Array Camera and Spectrograph (PACS) for the Herschel Space Observatory¹² and again with T-ReCS are consistent with the WISE data (Fig. 1; note especially the 2012 T-ReCS data), thus indicating that the mid-infrared emission from the dust orbiting this star has been consistently depleted to barely detectable levels since at least early 2010.

To determine the age of TYC 8241 2652 1, we obtained high-resolution optical spectra over four epochs from February 2008 to January 2009 with an echelle spectrograph mounted on the Siding Spring Observatory 2.3-m telescope. From these optical spectra, we estimate the age of the system from the lithium content in the stellar photosphere, Galactic space motion and rotational broadening of

absorption lines; details can be found in Supplementary Information. We adopt an age of ~ 10 Myr for TYC 8241 2652 1.

An important ingredient in understanding the vanishing mid-infrared emission from the dust orbiting TYC 8241 2652 1 is the initial state of the disk system. Given an age of ~ 10 Myr, the star could have been host either to an accreting protoplanetary disk rich in gas and dust or to a second-generation debris disk formed from the collisions of rocky objects orbiting the star¹³. The absence of strong Balmer H α emission from our optical spectroscopic measurements indicates that the star was not undergoing accretion of hydrogen-rich material at any significant level¹⁴ (see also Supplementary Information), and thus it is unlikely that such material was being transported inwards to the star as would be expected in a system with an active protoplanetary accretion disk. Another argument against TYC 8241 2652 1 having a protoplanetary accretion disk in the two decades before 2009 lies with the Herschel/PACS measurements. The sensitive upper limits in the far-infrared robustly rule out the presence of a substantial reservoir of cold disk material typical of those seen in protoplanetary disks. We thus conclude that the dusty material orbiting TYC 8241 2652 1 is the result of the collisions of rocky objects.

Table 1 | Parameters of TYC 8241 2652 1

Parameter	Value
Right ascension	12 h 9 min 2.25 s
Declination	$-51^\circ 20' 41.0''$
Galactic longitude	296.2104°
Galactic latitude	$+10.9728^\circ$
Visual magnitude	11.5 mag
Spectral type	K2 ± 1
Effective temperature	4950 ± 150 K
Proper motion in right ascension	-34.1 ± 2.1 mas yr ⁻¹
Proper motion in declination	-9.4 ± 2.0 mas yr ⁻¹
Heliocentric radial velocity	15 ± 1 km s ⁻¹
Lithium 6,708-Å EW	370 ± 10 mÅ
Balmer H α EW	0.0 ± 0.1 Å
Ca II K emission core EW	4.5 ± 0.5 Å
$v \sin(i)$	10 ± 1 km s ⁻¹
Distance from Earth	140 ± 20 pc (456 light yr)
Galactic space motions	-12 km s ⁻¹ (U), -24 km s ⁻¹ (V), -7 km s ⁻¹ (W)
Age	~ 10 Myr

J2000 equinox right ascension and declination are from the Two Micron All Sky Survey (2MASS) catalogue. Galactic longitude and latitude are derived from the 2MASS right ascension and declination. The spectral type and effective temperature are determined from line ratios²⁸ in the echelle spectra from the Siding Spring Observatory. Proper motion measurements are from the Tycho-2 catalogue⁵. The radial velocity is measured from our Siding Spring echelle spectra by cross-correlating a target spectrum with a standard star spectrum of known radial velocity. Four epochs of radial velocity measurements (14 February 2008, 14 June 2008, 13 July 2008 and 12 January 2009 UT) show no evidence for radial velocity variability within the measured errors (~ 1 – 2 km s⁻¹), ruling out any short-orbital-period stellar companions to TYC 8241 2652 1. The radial velocity quoted in the table is the average of the four separate measurements. The listed Balmer H α , Ca II K core reversal emission and lithium 6,708-Å equivalent widths (EWs) are averages over the four Siding Spring echelle epochs, and the uncertainty quoted is the standard deviation of those measurements. The stellar rotational velocity ($v \sin(i)$), where i is the angle of inclination of the stellar spin axis with respect to the line of sight towards Earth) was measured from the full-width at half-maximum (FWHM) depth of single absorption lines in the Siding Spring echelle spectra (which have an intrinsic resolution-element FWHM of ~ 13 km s⁻¹, a value that we subtract in quadrature from the FWHM measured in the spectra). Velocities of the TYC 8241 2652 1 system (relative to the Sun) towards the centre of the Galaxy, around the Galactic Centre and perpendicular to the Galactic plane (U , V , W) are calculated from Tycho-2 proper motions, our estimated photometric distance and the optical-echelle-measured radial velocity. Uncertainties in these values are roughly 2 km s⁻¹. See Supplementary Information for a discussion of the age.

¹Center for Astrophysics and Space Sciences, University of California, San Diego, California 92093-0424, USA. ²Department of Physics and Astronomy, University of California, Los Angeles, California 90095-1547, USA. ³Department of Physics and Astronomy, California State Polytechnic University, Pomona, Pomona, California 91768, USA. ⁴Department of Physics and Astronomy, University of Georgia, Athens, Georgia 30602, USA. ⁵Research School of Astronomy and Astrophysics, College of Mathematical and Physical Sciences, The Australian National University, Cotter Road, Weston Creek, Australian Capital Territory 2611, Australia.

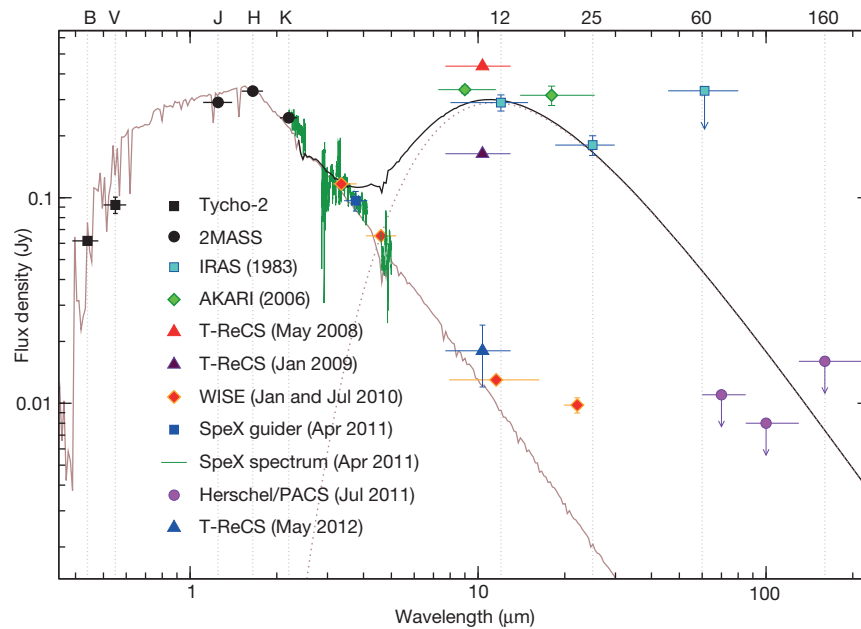


Figure 1 | Spectral energy distribution of TYC 8241 2652 1. Measurements and the associated epoch (for mid- and far-infrared data) are indicated in the legend. The solid brown curve is a synthetic stellar photosphere¹⁵ for a star with an effective temperature of 4,950 K that is fitted to the optical and near-infrared data. The dotted line is a black-body fit to the 12- and 25-μm IRAS excess data points. The temperature of this black body is 450 K and it suggests that roughly 11% of the optical and near-infrared starlight was being reprocessed into the mid-infrared by orbiting dust. The black solid line is the sum of the photosphere

and the black body at 450 K. Fitting a black body to the WISE and Herschel measurements suggests a dust temperature of roughly 200 K and a fractional infrared luminosity of 0.1%. Plotted flux density errors are 1 s.d. Some, for example those of the two earlier epochs of T-ReCS measurements, are smaller than the point sizes on the plot; for these measurements, the uncertainty is comparable to or less than 10% of the corresponding measurement. Horizontal lines through each data point represent the filter FWHM.

To estimate the dust temperature and the fractional infrared luminosity (L_{IR}/L_* , where L_* is the total stellar luminosity) of the dusty debris disk, we fit optical and near-infrared measurements out to the K_s band (2.1 μm) with a synthetic stellar atmosphere spectrum¹⁵ along with a black body at 450 K (Fig. 1) that models the pre-2009-epoch dust excess. Grains with a temperature of 450 K that are sufficiently large to radiate like black bodies at 10 μm and are situated in a disk optically thin to the stellar radiation field would orbit TYC 8241 2652 1 with a semi-major axis of ~ 0.4 AU. From the black-body fit, we find that $L_{\text{IR}}/L_* \approx 11\%$ (Fig. 1); such a value is significantly greater than those found previously for stars with warm debris disks¹⁶, but is less than that of the recently discovered, ~ 60 -Myr-old V488 Per system¹⁷. A geometrically thin, flat dust disk (such as Saturn's rings or some circumstellar debris disks) cannot absorb 11% of the luminosity of TYC 8241 2652 1 (ref. 18). To intercept such a large fraction of the incoming stellar light, the disk must be geometrically thick or otherwise deformed into a non-flat shape. Such a morphology could be suggestive of a substellar body that dynamically excites the dust particles, warps the

disk, or both^{19–21}. Fits to WISE and Herschel/PACS data allow a dust temperature only in the range $120 \text{ K} < T_{\text{dust}} < 250 \text{ K}$, indicating cool grains that orbit TYC 8241 2652 1 at a distance of ~ 2 AU, and a fractional infrared luminosity of $\sim 0.1\%$.

Given the luminosity of TYC 8241 2652 1 ($L_* \approx 0.7 L_\odot$, where L_\odot is the solar luminosity), grains with radii of ~ 0.2 μm and smaller will be radiatively ejected from the disk system. Roughly 5×10^{21} g of grains with radii of order 0.3 μm, which is slightly larger than the critical radius for radiative ejection, are required to produce the observed pre-2009-epoch infrared excess from TYC 8241 2652 1 (ref. 22). For a debris disk, the copious amounts of dust that were present suggest a system undergoing an active stage of terrestrial planet formation^{16,23}. The excess emission detected by WISE requires roughly 4–5 times less mass in cool, small (~ 0.3 -μm) dust grains than that estimated for the grains at 450 K detected closer to the star in the pre-2009 epoch. To have no WISE-detectable signature of grains at 450 K requires that such dust contribute less than 0.1% to the total fractional infrared luminosity in the post-2009-epoch measurements.

Table 2 | Mid-infrared flux measurements of TYC 8241 2652 1

Observation date (UT)	Instrument	Beam size (")	~ 10 -μm flux density (mJy)	~ 20 -μm flux density (mJy)
February–November 1983	IRAS	45 × 270 (PA = 132°)	309 ± 31	224 ± 22
May–November 2006	AKARI	5.5	335 ± 14	315 ± 34
6 May 2008	T-ReCS	0.4	436 ± 44	—
7 January 2009	T-ReCS	0.4	164 ± 16	—
8–10 January 2010	WISE	6.1	12.8 ± 0.4	9.4 ± 0.8
14–18 July 2010	WISE	6.1	12.3 ± 0.5	9.2 ± 1.0
1 May 2012	T-ReCS	0.4	18 ± 6	—

Central wavelengths for each instrument are 12 and 25 μm for the Infrared Astronomical Satellite (IRAS), 9 and 18 μm for AKARI, 10 μm for T-ReCS N-band imaging, and 11 and 22 μm for WISE. For each of the IRAS and WISE measurements, all available ancillary data products were examined to ensure reliability of the measured flux densities. T-ReCS observations were performed in clear, photometric conditions and were flux-calibrated using consecutive observations of stars with known mid-infrared fluxes. For instruments other than IRAS, the quoted beam size is the point spread function FWHM. The large, irregular IRAS beam size is a result of the focal plane detector mask used, and the position angle (PA) is the orientation of this rectangular mask on the sky when IRAS observed TYC 8241 2652 1. A PA of 0° is North and a PA of 90° is East. For each of the IRAS, AKARI and WISE measurements, the satellite-measured stellar position agrees with the Two Micron All Sky Survey position quoted in Table 1 to within the stated errors (after taking into account the stellar proper motion listed in Table 1). For T-ReCS measurements, the stellar position is not absolutely determined by the observations but instead is determined relative to the calibration star that is observed immediately before or after observations of TYC 8241 2652 1. On the T-ReCS detector array, the position of the mid-infrared source detected towards TYC 8241 2652 1 relative to the position of the calibration star is a reflection of how close the detected source position is to the input position, because the telescope slew precision is roughly 1" or better for small slews. We determine that, for each observation of TYC 8241 2652 1 with T-ReCS, the detected source lies within 1" of the stellar position in Table 1. We also note that there is only one source detected in the 28.8" × 21.6" T-ReCS field of view, and that each observation was sensitive enough to detect the photospheric flux level of TYC 8241 2652 1.

It is desirable to develop a physical model that can explain the observed disappearance of the disk of dust grains at 450 K. In Supplementary Information, we consider and reject models that rely on the disk material somehow being hidden from view, thus resulting in the diminished flux. In lieu of these models, we explore others in which the disk material is physically removed from its pre-2009-epoch location. If the number of grains with radius a that orbit the star follows a conventional $a^{-3.5}$ size distribution—and, hence, the fractional infrared luminosity scales like $a^{-0.5}$ (ref. 22)—we expect that removal of grains with radii up to ~ 1 mm would be required to eliminate the observational signature of dusty material orbiting at a separation of ~ 0.4 AU. For an $a^{-3.5}$ size distribution, the diminished mid-infrared flux requires that the total mass of dust particles with a less than ~ 1 mm located near 0.4 AU be smaller by a factor of ~ 100 . A steeper grain size distribution, with an exponent of -3.7 to -3.8 , would require removal of grains with radii up to ~ 100 μ m and would result in a reduction in the total grain mass by a factor of ~ 10 .

Of the models explored in Supplementary Information, only the collisional avalanche²⁴ and runaway accretion²⁵ models are potentially viable, although each has its problems (details regarding these models and their shortcomings can be found in Supplementary Information). It is worth noting that both models benefit if the steeper grain size distribution is assumed and that modelling of other stars with warm debris disks indicates that such a steep power-law slope may be present in such systems^{26,27}. Clear identification of a physical model that can reproduce the observations will require modelling specific to the case of TYC 8241 2652 1 and its continued observation. Although the exact circumstances are not yet clear, this system has clearly undergone a drastic event that promises to provide unique insight into the process by which rocky planets form.

Received 15 December 2011; accepted 4 May 2012.

1. Wetherill, G. W. Formation of the Earth. *Annu. Rev. Earth Planet. Sci.* **18**, 205–256 (1990).
2. Asphaug, E., Agnor, C. B. & Williams, Q. Hit-and-run planetary collisions. *Nature* **439**, 155–160 (2006).
3. Kenyon, S. J. & Bromley, B. C. Terrestrial planet formation. I. The transition from oligarchic growth to chaotic growth. *Astron. J.* **131**, 1837–1850 (2006).
4. Stewart, L. & Leinhardt, Z. Collisions between gravity-dominated bodies: II. The diversity of impact outcomes during the end stage of planet formation. *Astrophys. J.* **751**, 32 (2012).
5. Høg, E. et al. The Tycho-2 catalogue of the 2.5 million brightest stars. *Astron. Astrophys.* **355**, L27–L30 (2000).
6. Ishihara, D. et al. The AKARI/IRC mid-infrared all-sky survey. *Astron. Astrophys.* **514**, A1 (2010).
7. Wright, E. L. et al. The Wide-field Infrared Survey Explorer (WISE): mission description and initial on-orbit performance. *Astron. J.* **140**, 1868–1881 (2010).
8. Telesco, C. M. et al. GatorCam: the Gemini mid-infrared imager. *Proc. SPIE* **3354**, 534 (1998).
9. Rayner, J. T. et al. SpeX: a medium-resolution 0.8–5.5 micron spectrograph and imager for the NASA Infrared Telescope Facility. *Publ. Astron. Soc. Pacif.* **115**, 362–382 (2003).
10. Vacca, W. D., Cushing, M. C. & Rayner, J. T. A method of correcting near-infrared spectra for telluric absorption. *Publ. Astron. Soc. Pacif.* **115**, 389–409 (2003).
11. Cushing, M. C., Vacca, W. D. & Rayner, J. T. Spextool: a spectral extraction package for SpeX, a 0.8–5.5 micron cross-dispersed spectrograph. *Publ. Astron. Soc. Pacif.* **116**, 362–376 (2004).
12. Poglitsch, A. et al. The Photodetector Array Camera and Spectrometer (PACS) on the Herschel Space Observatory. *Astron. Astrophys.* **518**, L2 (2010).
13. Zuckerman, B. Dusty circumstellar disks. *Annu. Rev. Astron. Astrophys.* **39**, 549–580 (2001).
14. White, R. & Basri, G. Very Low mass stars and brown dwarfs in Taurus-Auriga. *Astrophys. J.* **582**, 1109–1122 (2003).
15. Hauschildt, P. H., Allard, F. & Baron, E. The NextGen model atmosphere grid for $3000 < T_{\text{eff}} < 10,000$ K. *Astrophys. J.* **512**, 377–385 (1999).
16. Melis, C., Zuckerman, B., Rhee, J. H. & Song, I. The age of the HD 15407 system and the epoch of final catastrophic mass accretion onto terrestrial planets around Sun-like stars. *Astrophys. J.* **717**, L57–L61 (2010).
17. Zuckerman, B. et al. Stellar membership and dusty debris disks in the alpha Persei Cluster. *Astrophys. J.* **752**, 58 (2012).
18. Jura, M. A tidally disrupted asteroid around the white dwarf G29–38. *Astrophys. J.* **584**, L91–L94 (2003).
19. Mouillet, D., Larwood, J. D., Papaloizou, J. C. B. & Lagrange, A. M. A planet on an inclined orbit as an explanation of the warp in the β Pictoris disc. *Mon. Not. R. Astron. Soc.* **292**, 896–904 (1997).
20. Akeson, R. et al. The circumbinary disk of HD 98800B: evidence for disk warping. *Astrophys. J.* **670**, 1240–1246 (2007).
21. Boley, A. et al. Constraining the planetary system of Fomalhaut using high-resolution ALMA observations. *Astrophys. J.* **750**, L21 (2012).
22. Rhee, J. H., Song, I. & Zuckerman, B. Warm dust in the terrestrial planet zone of a Sun-like Pleiades star: collisions between planetary embryos? *Astrophys. J.* **675**, 777–783 (2008).
23. Kenyon, S. J. & Bromley, B. C. Prospects for detection of catastrophic collisions in debris disks. *Astron. J.* **130**, 269–279 (2005).
24. Grigorieva, A., Artymowicz, P. & Thébault, Ph. Collisional dust avalanches in debris discs. *Astron. Astrophys.* **461**, 537–549 (2007).
25. Metzger, B., Rafikov, R. & Bochkarev, K. Global models of runaway accretion in white dwarf debris disks. *Mon. Not. R. Astron. Soc.* **423**, 505–528 (2012).
26. Lisse, C. M. et al. Abundant circumstellar silica dust and SiO gas created by a giant hypervelocity collision in the ~ 12 Myr HD172555 system. *Astrophys. J.* **701**, 2019–2032 (2009).
27. Currie, T. et al. Spitzer Infrared Spectrograph spectroscopy of the 10 Myr old EF Cha debris disk: evidence for phyllosilicate-rich dust in the terrestrial zone. *Astrophys. J.* **734**, 115 (2011).
28. Padgett, D. L. Atmospheric parameters and iron abundances of low-mass pre-main-sequence stars in nearby star formation regions. *Astrophys. J.* **471**, 847–866 (1996).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank J. Kastner for advice regarding X-ray data and M. Jura for suggesting the runaway accretion model. This work was based on observations obtained at the Gemini Observatory and makes use of data products from the Two Micron All Sky Survey and information from the SIMBAD and VizieR databases. C.M. acknowledges support from a LLNL Minigrant to UCLA and from the US National Science Foundation. This work was supported in part by NASA grants to UCLA and the University of Georgia.

Author Contributions The authors contributed equally to this work.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to C.M. (cmelis@ucsd.edu).

Optical nano-imaging of gate-tunable graphene plasmons

Jianing Chen^{1,2*}, Michela Badioli^{3*}, Pablo Alonso-González^{1*}, Sukosin Thongrattanasiri^{4*}, Florian Huth^{1,5*}, Johann Osmond³, Marko Spasenović³, Alba Centeno⁶, Amaia Pesquera⁶, Philippe Godignon⁷, Amaia Zurutuza Elorza⁶, Nicolas Camara⁸, F. Javier García de Abajo⁴, Rainer Hillenbrand^{1,9} & Frank H. L. Koppens³

The ability to manipulate optical fields and the energy flow of light is central to modern information and communication technologies, as well as quantum information processing schemes. However, because photons do not possess charge, a way of controlling them efficiently by electrical means has so far proved elusive. A promising way to achieve electric control of light could be through plasmon polaritons—coupled excitations of photons and charge carriers—in graphene^{1–5}. In this two-dimensional sheet of carbon atoms⁶, it is expected that plasmon polaritons and their associated optical fields can readily be tuned electrically by varying the graphene carrier density. Although evidence of optical graphene plasmon resonances has recently been obtained spectroscopically^{7,8}, no experiments so far have directly resolved propagating plasmons in real space. Here we launch and detect propagating optical plasmons in tapered graphene nanostructures using near-field scattering microscopy with infrared excitation light^{9–11}. We provide real-space images of plasmon fields, and find that the extracted plasmon wavelength is very short—more than 40 times smaller than the wavelength of illumination. We exploit this strong optical field confinement to turn a graphene nanostructure into a tunable resonant plasmonic cavity with extremely small mode volume. The cavity resonance is controlled *in situ* by gating the graphene, and in particular, complete switching on and off of the plasmon modes is demonstrated, thus paving the way towards graphene-based optical transistors. This successful alliance between nanoelectronics and nano-optics enables the development of active subwavelength-scale optics and a plethora of nano-optoelectronic devices and functionalities, such as tunable metamaterials¹², nanoscale optical processing, and strongly enhanced light–matter interactions for quantum devices¹³ and biosensing applications.

Surface plasmons are electromagnetic waves that propagate along the surface of a metal. Similar propagating waves are expected for graphene⁴. In fact, owing to the two-dimensional nature of the collective excitations in this material, the confinement of graphene plasmons is expected to be much stronger than that of metallic surface plasmons. However, launching and detecting graphene plasmons has so far remained a challenge: because of the large wavevector mismatch of graphene plasmons compared to free-space photons, plasmon excitation and detection by light is very inefficient. The first reports of graphene plasmon resonances were based on electron spectroscopies (inelastic electron scattering^{14–16} and photoemission spectroscopy¹⁷), used to spectrally probe excitation modes in graphene. Recently, resonant coupling of propagating terahertz waves to plasmons in micro-ribbons⁷ has been demonstrated, and infrared near-field microscopy has been applied to observe the coupling of graphene plasmons to phonons⁸. These pioneering works have revealed the interaction between low-energy photons and graphene plasmons in the spectral

domain. However, high-resolution nanoscale real-space imaging of the plasmonic modes is of fundamental importance to conclusive unveiling of propagating and localized plasmons in graphene sheets and nanostructures.

Here we visualize propagating and localized graphene plasmons in real space by scattering-type scanning near-field optical microscopy^{18,19} (scattering-type SNOM). Our scattering-type SNOM comprises an atomic force microscope (AFM) in which the metallized tip is illuminated with a focused infrared laser beam (Fig. 1a). The backscattered radiation is recorded simultaneously with the topography, yielding nanoscale resolved infrared near-field images (Methods). A representative near-field image is shown in Fig. 1b, where the tip is scanned over a tapered graphene ribbon on the carbon-terminated surface of 6H-SiC (ref. 20), illuminated by light of free-space wavelength $\lambda_0 = 9.7 \mu\text{m}$. One of the most distinct features in this image is the presence of fringes parallel to the edge of the ribbon in its wider part. The distance between fringe maxima is approximately constant at $\sim 130 \text{ nm}$ inside the ribbon. We interpret these fringes as follows: the near field at the tip apex locally excites radial surface waves that propagate along the surface and reflect at the edges, partially reaching

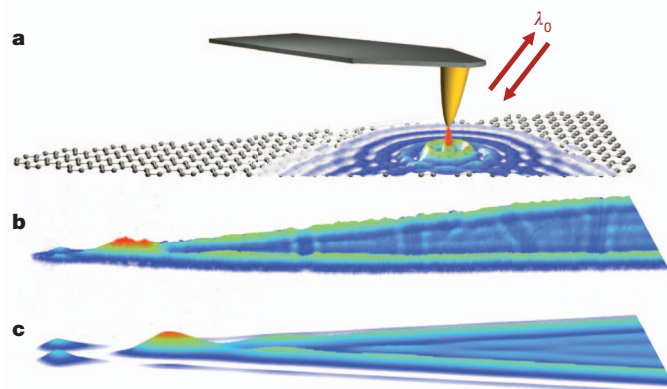


Figure 1 | Imaging propagating and localized graphene plasmons by scattering-type SNOM. **a**, Diagram of the experimental configuration used to launch and detect propagating surface waves in graphene (represented as blue rings). The metallized AFM tip (shown in yellow) is illuminated by an infrared laser beam with wavelength λ_0 . **b**, Near-field amplitude image acquired for a tapered graphene ribbon on top of 6H-SiC. The imaging wavelength is $\lambda_0 = 9.7 \mu\text{m}$. The tapered ribbon is $12 \mu\text{m}$ long and up to $1 \mu\text{m}$ wide. **c**, Colour-scale image of the calculated local density of optical states (LDOS) at a distance of 60 nm from the graphene surface, and assuming substrate $\epsilon_r = 1$. Simulation fitting parameters: graphene mobility $\mu = 1,000 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ and Fermi energy $E_F = 0.4 \text{ eV}$.

¹CIC nanoGUNE Consolider, 20018 Donostia-San Sebastián, Spain. ²Centro de Física de Materiales (CSIC-UPV/EHU) and Donostia International Physics Center (DIPC), 20018 Donostia-San Sebastián, Spain. ³ICFO-Institut de Ciències Fotoniques, Mediterranean Technology Park, 08860 Castelldefels, Barcelona, Spain. ⁴IQFR-CSIC, Serrano 119, 28006 Madrid, Spain. ⁵Neaspec GmbH, 82152 Martinsried, Munich, Germany. ⁶Graphenea SA, 20018 Donostia-San Sebastián, Spain. ⁷CNM-IMB-CSIC-Campus UAB, 08193 Bellaterra, Barcelona, Spain. ⁸GREMAN, UMR 7347, Université de Tours/CNRS, 37071 Tours Cedex 2, France. ⁹IKERBASQUE, Basque Foundation for Science, 48011 Bilbao, Spain.

*These authors contributed equally to this work.

the tip again. By recording the local field scattered by the tip, we probe the interference of forward- and backward-propagating plasmons. Within this basic physical picture, the maxima are separated by half the plasmon wavelength, $\lambda_p/2$. Thus, we experimentally find a plasmon wavelength $\lambda_p = 260$ nm, which is about a factor of 40 smaller than the free-space excitation wavelength. As we discuss in Methods, the near-field images can be interpreted in terms of the local density of optical states (LDOS). The calculated LDOS for a tapered ribbon is shown in Fig. 1c, matching very closely the experimental results (Fig. 1b), including the features at the narrower part of the ribbon.

Our observation of a remarkably strong reduction in the plasmon wavelength, $\lambda_p \approx \lambda_0/40$, can directly be attributed to the two-dimensionality and the unique conductance properties of graphene. Namely, the plasmonic properties of graphene are related to its optical conductivity^{1,21}, σ (for example, $2\pi/\lambda_p \approx (\epsilon_r + 1)\omega/(4\pi\text{Im}\{\sigma\})$, where ϵ_r is the substrate permittivity and ω is the frequency). For sufficiently high doping, quantified through a Fermi energy E_F exceeding the plasmon energy E_p , this yields^{4,13}:

$$\lambda_p \approx \lambda_0 \alpha \frac{E_F}{E_p \epsilon_r + 1} \quad (1)$$

We note that this simplified equation reveals a relation between the plasmon wavelength and the free-space wavelength that is governed by the fine-structure constant, $\alpha \approx 1/137$. The observed $\lambda_p = 260$ nm is in reasonable agreement with the theoretical prediction of equation (1) for the specific substrate (ϵ_r for SiC is 1.9 for $\lambda_0 = 9.7$ μm), which yields $\lambda_p = 305$ nm assuming $E_F \approx 0.4$ eV. This value is about a factor of two higher than the intrinsic substrate-induced doping found in earlier studies of graphene on the carbon-terminated surface of 6H-SiC (ref. 22). We speculate that narrow ribbons exhibit larger carrier densities, and this will be addressed in future studies.

In Fig. 2, we present a more detailed experimental study of the plasmon properties in graphene nanostructures by taking advantage

of the strong dependence of the dielectric constant of the SiC substrate, $\epsilon_{r,\text{SiC}}$, on the excitation wavelength²³. This allows us to tune the plasmon wavelength over a wide spectral range by just slightly changing the excitation wavelength, as the plasmon wavelength depends strongly on the dielectric constant of the substrate. The near-field images of relatively wide ribbons are displayed in Fig. 2a, showing that the spacing of the interference fringes decreases considerably with increasing $\epsilon_{r,\text{SiC}}$. This observation is qualitatively consistent with equation (1), as a larger substrate permittivity yields a smaller graphene plasmon wavelength. Quantitatively, we obtain good agreement between the plasmon wavelengths extracted from the near-field images (Fig. 3a, symbols) and the prediction of equation (1) for graphene on SiC (Fig. 3a, solid curves), using literature values for the dielectric constant of SiC (ref. 23) and an intrinsic doping $E_F = 0.4$ eV.

Our experimental observation of an extremely short plasmon wavelength compared to the excitation wavelength is associated with an extraordinary confinement of the infrared field perpendicular to the graphene sheet, characterized by a decay length $\delta \approx \lambda_p/2\pi$ (ref. 11). This means that narrow graphene ribbons are ideally suited to confine light down to extremely small volumes. In Fig. 2b, we show near-field images of the tapered ribbons where the width W reaches values smaller than the plasmon wavelength λ_p . These images clearly reveal two distinct localized modes (indicated by red and white arrows) that coexist with a resonant enhancement of the near-field signal, comparable to the observations in ref. 24. The resonance condition depends on λ_p and the ribbon width W , as we observe a clear shift of the localized modes to a wider part of the ribbons for increasing λ_p . The width W for which these two modes occur, normalized to the plasmon wavelength λ_p , is shown in Fig. 3b, from which we extract the resonance conditions $W \approx 0.3\lambda_p$ and $\sim 0.6\lambda_p$.

To better understand the physical mechanisms that underlie the observation of these resonant optical modes, as well as the interference

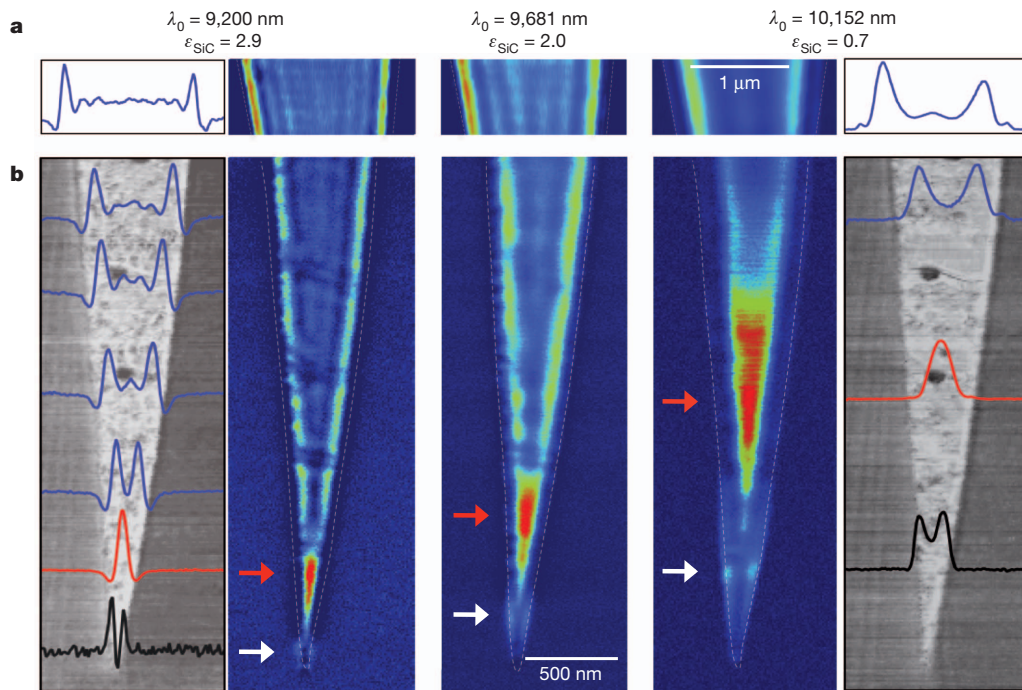


Figure 2 | Controlling the plasmon wavelength over a wide range.

a, b, Coloured plots show near-field optical images taken with imaging wavelengths (λ_0) of 9,200 nm (left), 9,681 nm (middle) and 10,152 nm (right), corresponding respectively to SiC dielectric constants of 2.9, 2.0 and 0.7. **a,** Images of a graphene ribbon ~ 1 μm wide, revealing a strong dependence of the fringe spacing, and thus plasmon wavelength, on the excitation wavelength;

b, images of a tapered graphene ribbon; both ribbons are on the same 6H-SiC substrate. The topography (obtained by AFM) is shown in greyscale in the leftmost and rightmost panels, and outlined by dashed lines in the central, coloured panels. The line traces in the leftmost and rightmost panels are extracted from the near-field images for $\lambda_0 = 9,200$ nm and $\lambda_0 = 10,152$ nm. Red and white arrows indicate the resonant localized modes.

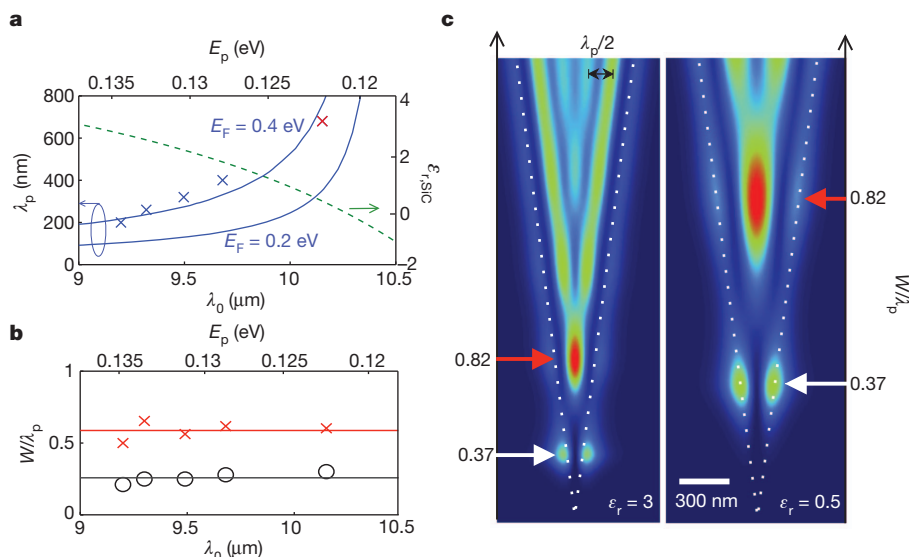


Figure 3 | Comparison of theoretical model with experimental results. **a**, Experimentally extracted plasmon wavelength λ_p as a function of incident wavelength λ_0 . Values for λ_p are obtained from interference fringes (blue crosses) and localized modes (red crosses), compared to the calculated plasmon dispersion (blue curves, see Supplementary Information) for graphene assuming intrinsic doping of 0.2 and 0.4 eV on a SiC-6H substrate. Green dashed line, SiC substrate permittivity. **b**, Experimentally obtained resonance

fringes, we calculate maps of the local density of optical states (LDOS) of the graphene ribbons. The LDOS maps calculated for two different values of the substrate permittivity $\epsilon_{r,\text{SiC}}$ are shown in Figs 1c and 3c. As in the experimental scattering-type SNOM images, the LDOS maps reveal interference fringes parallel to the ribbon edge, and localized modes near the tip of the ribbon. The fringe spacing matches quantitatively the experimental results and the spacing increases with decreasing ϵ_r , associated with an increase in λ_p , as predicted by equation (1). The good agreement between experiment and theory confirms that the fringes in the wider part of the ribbon are due to plasmon interference caused by plasmon reflections at the graphene edges. We remark that both the LDOS and the experimental images exhibit their maximum away from the graphene edge, and that the fringe spacing increases slightly closer to the edge. This can be explained by the electromagnetic boundary conditions at the edges (further discussed below) and the fact that the plasmon wavevector perpendicular to the edge does not have a single value, but rather a finite distribution around $2\pi/\lambda_p$ (see Supplementary Information).

The comparison between the calculated LDOS maps and the experimental data in Fig. 2 can be used to estimate plasmon propagation distances. We observe five well-defined interference fringes away from a single edge. The peak close to the edge is relatively strong, which we attribute to enhanced near-field coupling between tip and graphene close to the edge, and multiple plasmon reflections between the metallic tip and the edge. The fringes inside the ribbon decay, owing to both the circular character of the plasmons and intrinsic losses. These observations are consistent with our LDOS calculations for plasmon losses corresponding to a mobility of $1,200 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ (ref. 4). This value was obtained by electrical characterization of similar graphene ribbons under ambient experimental conditions. In particular, the observed coexistence of a strong reduction in plasmon wavelength (and thus strong optical field confinement) and relatively long propagation distance is very promising, and a unique feature of plasmons carried by graphene. We emphasize that much longer propagation distances are expected for higher-mobility graphene.

In our LDOS model interpretation, the localized modes near the tip of the graphene ribbon (marked by arrows in Fig. 2b) are explained as localized graphene plasmon resonances, which occur for specific

conditions W/λ_p extracted from localized-mode measurements. Red crosses and black circles correspond to the modes indicated by red and white arrows in Fig. 2, respectively. **c**, Spatial distribution of the LDOS calculated for homogeneous ribbons of increasing width (from bottom to top), supported on a dielectric with $\epsilon_r = 3$ (left) or $\epsilon_r = 0.5$ (right). The ribbon width of the two lowest-order modes is shown in units of the plasmon wavelength of extended graphene, λ_p .

values of the ribbon width ($W = 0.37\lambda_p$ and $0.82\lambda_p$)²⁵, where the strong concentration of the electromagnetic field yields an enhanced plasmon-dipole interaction¹³, and therefore, an increase in the near-field signal. For both theory and experiment, the profiles of the two localized modes are distinctly different from those of conventional Fabry–Perot cavity modes. For example, the lowest-order mode (indicated by white arrows in Fig. 2b) exhibits field maxima at the graphene edges, whereas for a conventional lowest-order Fabry–Perot mode the field is maximum in the middle. This is because graphene plasmons are being reflected at the boundaries with a reflection coefficient of approximately one (zero phase), rather than the coefficient of minus one (π phase) characteristic of the conventional Fabry–Perot model (see Supplementary Information).

One of the most appealing advantages of graphene plasmonics is the capability to control and switch nanoscale optical fields *in situ*. Here we demonstrate very effective electrical control of nanoscale optical fields by applying an electric field perpendicular to the graphene sheet, which allows us to vary the carrier density in the ribbon. To this end, we have fabricated tapered ribbons based on CVD-grown graphene on a SiO_2 substrate with a Si backgate. By applying a backgate voltage V_B , we tune the carrier density and thus the Fermi energy $E_F \approx (V_B - V_D)^{1/2}$, where V_D is the voltage that needs to be applied to offset the intrinsic doping (that is, to reach the Dirac point; V_D is extracted from optical measurements, as we discuss below). The effect of changing V_B on the near-field images is shown in Fig. 4a, where the Fermi energy is tuned over a wide range, from about 0 to 0.15 eV. For $V_B - V_D > 10 \text{ V}$, the general near-field features are comparable to those of ribbons on SiC substrates, including the two local ribbon resonances indicated by white and red arrows in Fig. 4a. By increasing V_B , we find that the resonances (signal maxima) shift towards larger ribbon width, which we attribute to an increase in plasmon wavelength when the carrier density, and thus also the Fermi energy, increases (see equation (1)). The extracted value of λ_p as a function of gate voltage is shown in the lower part of Fig. 4b (red circles correspond to the tapered ribbon shown in Fig. 4a, and green crosses correspond to additional ribbons shown in Supplementary Information). The calculated plasmon dispersion, represented by the blue contour plot in the bottom part of Fig. 4b, is based on the random-phase approximation, which includes

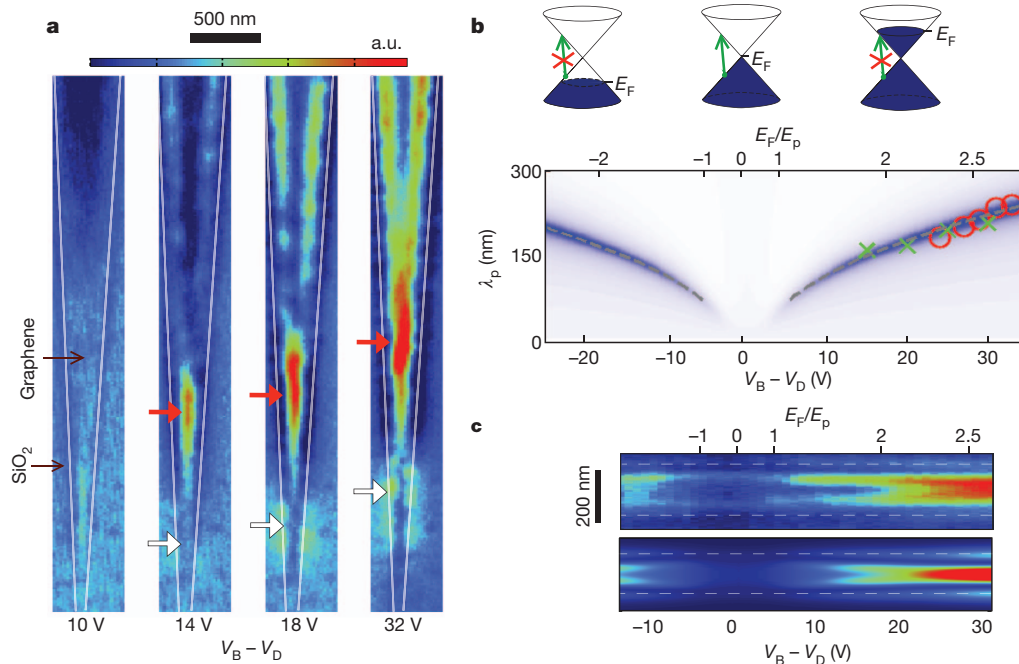


Figure 4 | Plasmonic switching and active control of the plasmon wavelength by electrical gating. **a**, Near-field amplitude images for tapered (CVD-grown, obtained from Graphenea SA) graphene ribbons on a Si/SiO₂ (300 nm) substrate, acquired while applying backgate voltages V_B ranging from -15 to $+11$ V. We extract the Dirac voltage V_D from optical images and by fitting the data to the model. Localized modes are indicated by white and red arrows. The illumination wavelength is $\lambda_0 = 11.06$ μm . **b**, Bottom panel, plasmon wavelength experimentally extracted from localized mode resonances indicated by red arrows in **a**. Red circles represent the data sets presented in **a**, while green markers correspond to one additional data set presented in Supplementary Information. Top panel, diagrams showing the Dirac cone of

plasmon damping by interband excitations for small carrier densities ($|E_F| \lesssim E_p$), as illustrated by the diagrams in the top part of Fig. 4b. These excitations can strongly damp the plasmons, but to first order these transitions are suppressed for $|E_F| \gtrsim E_p$. The data agree well with the calculated plasmon dispersion, taking into account an electrostatic enhancement of the carrier concentration due to the narrow width of the ribbon compared to the thickness of the oxide (see Supplementary Information).

The effect of plasmon damping offers the intriguing capability to actively switch graphene plasmons on and off by electric fields. Experimentally, we clearly observe very strong plasmon damping in the left panel of Fig. 4a (corresponding to $E_F \lesssim E_p$), where the ribbon does not show any signal compared to the substrate. We illustrate electrostatic switching of graphene plasmons in more detail in Fig. 4c, which shows line scans across a ribbon of width $W = 200$ nm (vertical axis), while changing V_B (horizontal axis). At the Dirac point ($V_B = V_D$), the near-field signal is dramatically depleted on the whole ribbon. With increasing Fermi energy, plasmon modes emerge at both sides of the Dirac point. These modes are attributed to plasmons carried by either p- or n-type charge carriers. The calculated LDOS profiles as a function of Fermi energy (lower plot of Fig. 4c) is in excellent agreement with the experimental observations.

Here and in ref. 24, electrical control of confined and propagating plasmons is demonstrated, thus providing a solution to a major problem in plasmonics, as it facilitates the design and miniaturization of active nanoscale photonic devices^{26,27}. This leads to a new paradigm in optical and opto-electronic telecommunications and information processing. As an alternative to plasmon excitation and detection by (effective) dipoles, plasmons can also be resonantly excited by light in graphene nanocavities¹³, enabling strong enhancement of light

absorption in graphene²⁸, and a new basis for infrared detectors and light-harvesting devices.

METHODS SUMMARY

Plasmon nano-imaging. The scattering-type SNOM used for this work (from Neaspec GmbH) employs metal AFM tips as near-field probes, which are illuminated by infrared light from a grating-tunable CO₂ laser. The tip acts as an optical antenna that converts the incident light into a localized near field below the tip apex¹⁸. The nanoscale field concentration provides the required momentum^{11,29} for launching plasmons on graphene, as illustrated in Fig. 1a. Plasmon reflection at the graphene edges produces plasmon interference, which is imaged by recording the light elastically scattered by the tip with a pseudo-heterodyne interferometer¹⁹. In order to suppress background scattering from the tip shaft and the sample, the tip is vibrated vertically with 50-nm amplitude at a frequency of about $\Omega = 300$ kHz, and the detector signal is demodulated at a higher harmonic $n\Omega$. With this technique, we record the amplitude and phase of the scattered field (see Supplementary Information). The amplitude is shown in the near-field images in Figs 1, 2 and 4.

LDOS calculations. In order to understand the near-field images, we model the microscope tip as a vertically oriented point dipole¹¹, which is scanned 60 nm above the graphene. The dipole launches plasmons^{25,30} that are reflected at the ribbon edges. These plasmons act back on the tip, and are subsequently scattered into photons, which we detect. The detected signal is strongly correlated with the vertical component of the LDOS. We calculate the LDOS by solving the Maxwell equations for a dipole source \mathbf{p} at location \mathbf{r}_0 :

$$\text{LDOS} = \text{LDOS}_{\text{vac}} + \frac{1}{2\pi^2\omega|\mathbf{p}|^2} \text{Im}\{\mathbf{E}^{\text{ref}}(\mathbf{r}_0) \cdot \mathbf{p}^*\}$$

where \mathbf{E}^{ref} is the field reflected by nearby structures and evaluated at the position of the source dipole. This is the procedure we actually follow to obtain the LDOS in this work, and \mathbf{E}^{ref} is calculated by means of the boundary-element method for a dipole source. In order to simulate two-dimensional LDOS maps for a dipole 60 nm above a tapered graphene ribbon, we combine one-dimensional LDOS profiles of graphene ribbons of fixed width. We justify this approach on the

grounds that the ribbon width along the long axis of the graphene triangle varies adiabatically, yielding a weak reflection at the tip of the ribbon. In addition, the plasmon propagation distance is of the order of a few plasmon wavelengths. Owing to these two effects, the intensity of back-reflected plasmons from the tip of the ribbon is expected to be weak. This is consistent with the absence of modulations in the experimental near-field images.

Received 1 March; accepted 21 May 2012.

Published online 20 June 2012.

1. Wunsch, B., Stauber, T., Sols, F. & Guinea, F. Dynamical polarization of graphene at finite doping. *N. J. Phys.* **8**, 318 (2006).
2. Hwang, E. H. & das Sarma, S. Dielectric function, screening, and plasmons in two-dimensional graphene. *Phys. Rev. B* **75**, 205418 (2007).
3. Polini, M. *et al.* Plasmons and the spectral function of graphene. *Phys. Rev. B* **77**, 081411(R) (2008).
4. Jablan, M., Buljan, H. & Soljačić, M. Plasmonics in graphene at infrared frequencies. *Phys. Rev. B* **80**, 245435 (2009).
5. Hill, A., Mikhailov, S. A. & Ziegler, K. Dielectric function and plasmons in graphene. *Europhys. Lett.* **87**, 27005 (2009).
6. Novoselov, K. S. *et al.* Electric field effect in atomically thin carbon films. *Science* **306**, 666–669 (2004).
7. Ju, L. *et al.* Graphene plasmonics for tunable terahertz metamaterials. *Nature Nanotechnol.* **6**, 630–634 (2011).
8. Fei, Z. *et al.* Infrared nanoscopy of Dirac plasmons at the graphene–SiO₂ interface. *Nano Lett.* **11**, 4701–4705 (2011).
9. Keilmann, F. & Hillenbrand, R. Near-field microscopy by elastic light scattering from a tip. *Phil. Trans. A* **362**, 787–805 (2004).
10. Huber, A., Ocelic, N., Kazantsev, D. & Hillenbrand, R. Near-field imaging of mid-infrared surface phonon polariton propagation. *Appl. Phys. Lett.* **87**, 081103 (2005).
11. Novotny, L. & Hecht, B. *Principles of Nano-optics* (Cambridge Univ. Press, 2006).
12. Vakil, A. & Engheta, N. Transformation optics using graphene. *Science* **332**, 1291–1294 (2011).
13. Koppens, F. H. L., Chang, D. E. & García de Abajo, F. J. Graphene plasmonics: a platform for strong light–matter interactions. *Nano Lett.* **11**, 3370–3377 (2011).
14. Liu, Y., Willis, R., Emtsev, K. & Seyller, T. Plasmon dispersion and damping in electrically isolated two-dimensional charge sheets. *Phys. Rev. B* **78**, 201403(R) (2008).
15. Eberlein, T. *et al.* Plasmon spectroscopy of free-standing graphene films. *Phys. Rev. B* **77**, 233406 (2008).
16. Zhou, W. *et al.* Atomically localized plasmon enhancement in monolayer graphene. *Nature Nanotechnol.* **7**, 161–165 (2012).
17. Bostwick, A., Ohta, T., Seyller, T., Horn, K. & Rotenberg, E. Quasiparticle dynamics in graphene. *Nature Phys.* **3**, 36–40 (2007).
18. Hillenbrand, R., Taubner, T. & Keilmann, F. Phonon-enhanced light matter interaction at the nanometre scale. *Nature* **418**, 159–162 (2002).
19. Ocelic, N., Huber, A. & Hillenbrand, R. Pseudoheterodyne detection for background-free near-field spectroscopy. *Appl. Phys. Lett.* **89**, 101124 (2006).
20. Camara, N. *et al.* Current status of self-organized epitaxial graphene ribbons on the C face of 6H–SiC substrates. *J. Phys. D Appl. Phys.* **43**, 374011 (2010).
21. Castro Neto, A. H., Peres, N. M. R., Novoselov, K. S. & Geim, A. K. The electronic properties of graphene. *Rev. Mod. Phys.* **81**, 109–162 (2009).
22. Crassee, I. *et al.* Multicomponent magneto-optical conductivity of multilayer graphene on SiC. *Phys. Rev. B* **84**, 035103 (2011).
23. Hofmann, M., Zywiez, A., Karch, K. & Bechstedt, F. Lattice dynamics of SiC polytypes within the bond-charge model. *Phys. Rev. B* **50**, 13401 (1994).
24. Fei, Z. *et al.* Gate-tuning of graphene plasmons revealed by infrared nano-imaging. *Nature* <http://dx.doi.org/10.1038/nature11253> (this issue).
25. Christensen, J., Manjavacas, A., Thongrattanasiri, S., Koppens, F. H. L. & García de Abajo, F. J. Graphene plasmon waveguiding and hybridization in individual and paired nanoribbons. *ACS Nano* **6**, 431–440 (2012).
26. Atwater, H. A. The promise of plasmonics. *Sci. Am.* **296**, 56–62 (2007).
27. Zia, R., Schuller, J. A., Chandran, A. & Brongersma, M. L. Plasmonics: the next chip-scale technology. *Mater. Today* **9**, 20–27 (2006).
28. Thongrattanasiri, S., Koppens, F. & García de Abajo, F. J. Complete optical absorption in periodically patterned graphene. *Phys. Rev. Lett.* **108**, 047401 (2012).
29. Hecht, B., Bielefeldt, H., Novotny, L., Inoué, Y. & Pohl, D. W. Local excitation, scattering, and interference of surface plasmons. *Phys. Rev. Lett.* **77**, 1889–1892 (1996).
30. Nikitin, A. Y., Guinea, F., García-Vidal, F. J. & Martín-Moreno, L. Edge and waveguide terahertz surface plasmon modes in graphene microribbons. *Phys. Rev. B* **84**, 161407(R) (2011).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank L. Novotny, N. van Hulst, R. Quidant and P. Jarillo-Herrero for discussions. This work was supported in part by the Fundació Cellex Barcelona, the Spanish MICINN (MAT2010-14885 and Consolider NanoLight.es), the European FP7 projects FP7-HEALTH-F5-2009-241818-NANOANTENNA, FP7-ICT-2009-4-248909-LIMA and FP7-ICT-2009-4-248855-N4E, the ERC Starting grant no. 258461 (TERATOMO), and the ERC Career integration grant GRANOP.

Author Contributions J.C., P.A.-G., F.H., F.H.L.K. and R.H. carried out the near-field imaging experiments and participated in data analysis. M.S. participated in data analysis. S.T. and F.J.G.d.A. contributed to the interpretation of the data and developed analytical and computational theoretical tools. N.C., P.G., A.C., A.P. and A.Z.E. provided materials. M.B. and J.O. fabricated the devices. J.G.d.A., R.H. and F.H.L.K. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/nature. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to F.J.G.d.A. (J.G.deAbajo@csic.es; theory) or R.H. (r.hillenbrand@nanogune.eu; near-field microscopy) or F.H.L.K. (frank.koppens@icfo.es; graphene).

Biophysical mechanism of T-cell receptor triggering in a reconstituted system

John R. James¹ & Ronald D. Vale¹

A T-cell-mediated immune response is initiated by the T-cell receptor (TCR) interacting with peptide-bound major histocompatibility complex (pMHC) on an infected cell. The mechanism by which this interaction triggers intracellular phosphorylation of the TCR, which lacks a kinase domain, remains poorly understood. Here, we have introduced the TCR and associated signalling molecules into a non-immune cell and reconstituted ligand-specific signalling when these cells are conjugated with antigen-presenting cells. We show that signalling requires the differential segregation of a phosphatase and kinase in the plasma membrane. An artificial, chemically controlled receptor system generates the same effect as TCR-pMHC, demonstrating that the binding energy of an extracellular protein-protein interaction can drive the spatial segregation of membrane proteins without a transmembrane conformational change. This general mechanism may extend to other receptors that rely on extrinsic kinases, including, as we demonstrate, chimaeric antigen receptors being developed for cancer immunotherapy.

In addition to intercellular communication mediated by soluble molecules, two cells can transmit signals through membrane-associated receptors and ligands. Adaptive immunity represents such a system, in which the MHC protein on the surface of antigen-presenting cells (APCs) interacts with the TCR on T lymphocytes. If the TCR binds pMHC of the right complementarity, the interaction results in tyrosine phosphorylation of the TCR (herein referred to as TCR 'triggering') and the initiation of signals that activate the T cell¹. The TCR has no intrinsic kinase activity, unlike many other receptors², and instead relies on a T-cell-specific kinase called Lck (ref. 3). Also distinct from other systems, the phosphorylatable tyrosine residues of the TCR (the immunoreceptor tyrosine-based activation motifs (ITAMs))⁴ do not reside on the polypeptides that contact the pMHC (α , β) but instead are contained on tightly associated CD3 subunits (γ , δ , ϵ , ζ). The phosphorylated ITAMs then bind a second kinase, ZAP70, which is subsequently activated and drives downstream signalling⁵.

Despite considerable work, the mechanism by which pMHC binding leads to TCR triggering remains poorly understood (reviewed in ref. 6). Some models propose that pMHC binding evokes a conformational change in the TCR that makes its cytoplasmic ITAM domains more accessible to Lck kinase⁷. Alternative triggering hypotheses include activation through the aggregation of TCR molecules⁶, and 'kinetic segregation'⁸, where TCR phosphorylation is favoured by its partitioning into plasma membrane domains that contain Lck kinase but are depleted of CD45 (also known as PTPRC), an abundant transmembrane phosphatase. However, although TCR clustering⁹ and the segregation of CD45 away from the TCR have been observed¹⁰, it has not been established whether such events are necessary or sufficient for signal transduction across the plasma membrane. In addition, the physical basis of protein segregation within the plasma membrane is unclear.

Reconstitution of a biological phenomenon with defined components has proven to be a powerful means for dissecting molecular mechanisms. We have made use of this approach by introducing the genes encoding the TCR and other proteins required for regulating its phosphorylation into a non-immune cell and recapitulating

TCR triggering when this cell forms a conjugate with an APC. Because each protein can be introduced separately and is genetically engineered, this system has allowed us to test models of TCR triggering and the roles of individual proteins in a manner that is difficult to achieve with native T cells.

Reconstitution of regulated TCR triggering

We first sought to reconstitute Lck-mediated TCR phosphorylation in a non-immune cell and then determine which factors are needed to keep the TCR quiescent (Fig. 1a). As the basis of our reconstitution, we expressed¹¹ the complete set of protein chains of the 1G4 TCR (ref. 12) in the plasma membrane of HEK cells (hereafter referred to as HEK-1G4) (Supplementary Methods and Supplementary Fig. 1). The expressed TCR complex did not show detectable phosphorylation (assayed by a phospho-specific antibody to the CD3 ζ chain, an essential TCR subunit required for signalling^{13,14}) unless Lck and ZAP70 were co-expressed (Fig. 1b). Lck kinase activity, as detected by measuring levels of activating (Tyr 394) and inhibitory (Tyr 505) phosphorylation³, seemed to be unaffected by the presence of the TCR or ZAP70 (Fig. 1b). However, ZAP70 activity, as measured by increased Tyr 493 phosphorylation⁵, was only detectable in the presence of both Lck and the TCR (Fig. 1b), which is in agreement with previous data suggesting that this kinase is inactive until it binds to phosphorylated CD3 ζ ITAMs¹⁵ (Fig. 1a). We confirmed the activity of ZAP70 by demonstrating phosphorylation of co-expressed LAT, its downstream substrate and critical adaptor protein for T-cell signalling (Supplementary Fig. 2a).

To establish a quiescent system that could be activated by pMHC, we next sought to restrain the kinase activity of Lck. CSK induces an 'inactive' conformation of Lck by phosphorylating its carboxy terminus (Tyr 505) (ref. 16). However, co-expressing CSK and CSK-binding protein (CBP, also known as PAG1), which localizes CSK to the plasma membrane (Fig. 1a), was insufficient to repress Lck phosphorylation of CD3 ζ (Fig. 1c). CD45 is a tyrosine phosphatase that modulates T-cell signalling in a complex manner by dephosphorylating the inhibitory Tyr 505 and activating Tyr 394 of Lck (refs 17, 18) and the ITAM tyrosines of the TCR (Fig. 1a). Co-expression of CD45 with Lck severely

¹Howard Hughes Medical Institute and Department of Cellular and Molecular Pharmacology, University of California, San Francisco, 600 16th Street, San Francisco, California 94158, USA.

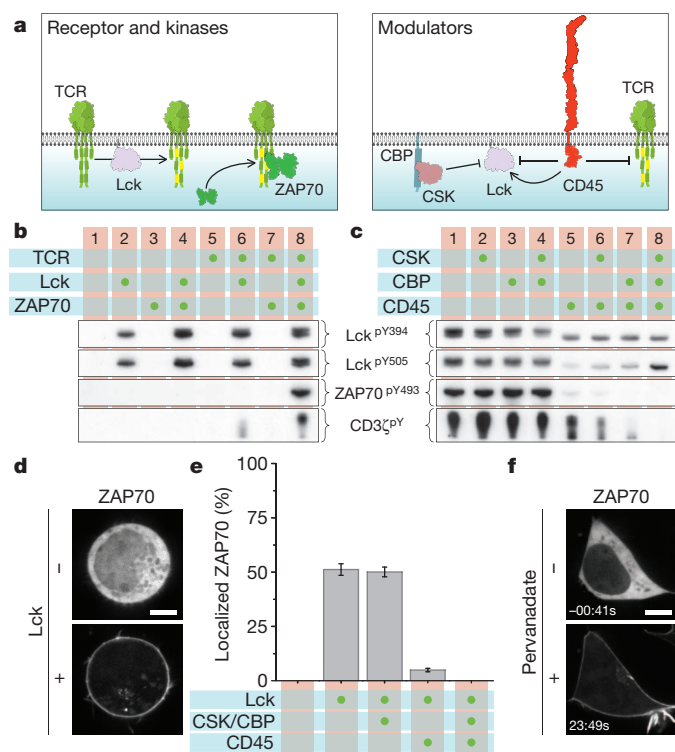


Figure 1 | Regulated TCR triggering in an engineered HEK cell line.

a, Schematic representation of molecules transfected into HEK cells. **b**, Western blot of phosphorylated proteins after transfection of HEK cells with selected molecules (green circles). pY, phosphorylated tyrosine. **c**, Cells transfected with the TCR, Lck and ZAP70 were transfected with additional molecules (green circles), showing the synergistic action of modulatory proteins CSK, CBP and CD45 to restrain Lck. The decreased phosphorylation between lanes five and seven is due to endogenous CSK recruitment. **d**, Confocal images of HEK-1G4 showing ZAP70-GFP recruitment to the plasma membrane in the presence of Lck. **e**, Quantification of ZAP70 relocalization with indicated molecules transfected in HEK-1G4 cells. Data are mean \pm s.e.m. of three independent experiments (~ 300 cells per experiment). **f**, Addition of 100 μ M pervanadate to HEK-1G4 cells (expressing components in lane eight of c) caused the accumulation of membrane-localized ZAP70. Scale bars, 5 μ m.

diminished Lck-induced ZAP70 activation but only modestly inhibited phosphorylation of the CD3 ζ chain of the TCR (Fig. 1c). However, simultaneous expression of CSK, CBP and CD45 considerably reduced CD3 ζ phosphorylation (Fig. 1c). This synergy depended on CBP (Fig. 1c), indicating that membrane recruitment of CSK is required for its potency. Thus, the two major activities known to repress TCR phosphorylation in T cells are sufficient, when combined, to keep the TCR in a quiescent state in this reconstituted cell system.

ZAP70 is normally cytosolic but binds to phosphorylated TCR ITAMs. The translocation of ZAP70-green fluorescent protein (GFP) from the cytosol to the plasma membrane therefore provides a microscopy-based assay of TCR triggering. Indeed, Lck expression in HEK-1G4 cells caused ZAP70-GFP to accumulate at the plasma membrane (Fig. 1d, e), whereas co-expression of CSK, CBP and CD45 delocalized ZAP70 to the cytoplasm (Fig. 1e). Furthermore, ZAP70-GFP translocated rapidly to the plasma membrane upon CD45 phosphatase inhibition by pervanadate (Fig. 1f and Supplementary Movie 1). Thus, this visual assay provides a dynamic readout of TCR triggering.

Reconstitution of a T cell-APC conjugate

Having identified a minimal set of components that could be expressed in HEK-1G4 cells to mimic the basal or 'off' state of their T-cell counterpart, we next attempted to trigger signalling by having the 1G4 TCR interact with its ligand, the MHC class-I complex bound

with a short peptide (ESO9V) (ref. 19) and expressed in the Raji B-cell line as the APC (Fig. 2a and Supplementary Methods). Because HEKs and Raji B cells have no intrinsic affinity for each other (Supplementary Fig. 3a), we expressed the immune-cell adhesion proteins CD2 and ICAM1 on the HEK-1G4 cells (Fig. 2a), which can interact with their respective counter-receptors CD58 and LFA-1 expressed endogenously on Raji cells (Supplementary Fig. 1). This strongly increased the conjugation between the two cell types, with the LFA-1-ICAM1 interaction being the predominant driver (Supplementary Fig. 3a).

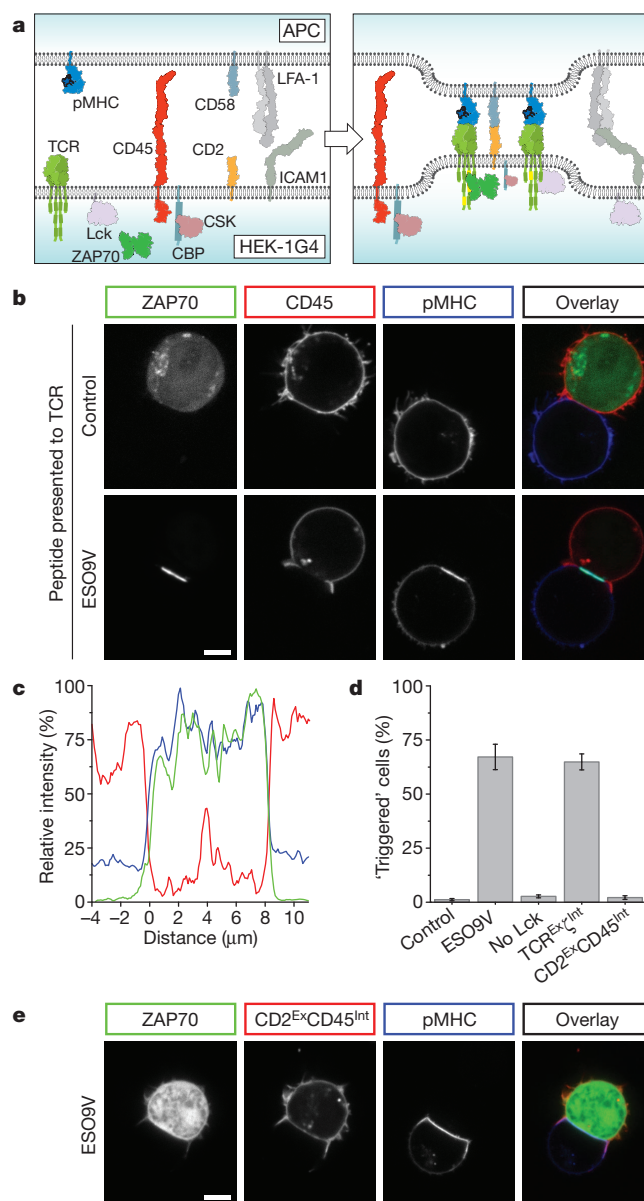


Figure 2 | The exclusion of CD45 phosphatase is necessary and sufficient for TCR triggering. **a**, Proteins expressed in HEK-1G4 for cell conjugation. **b**, HEK-1G4 cells, expressing all components shown in **a**, were conjugated with APCs (Raji cells) expressing cognate pMHC (ESO9V) or a control pMHC. Coloured boxes denote protein representation in the overlay image. **c**, A representative line profile of membrane fluorescence from CD45 (red), pMHC (blue) and ZAP70 (green) at the conjugate interface. **d**, Quantification of triggering (defined as unambiguous recruitment of ZAP70 to conjugate region) for all conjugates described in text. Data are mean \pm s.e.m. of independent experiments ($n = 4$ or 5, 30–150 conjugates per experiment). **e**, Forcing CD45 into the conjugate region by fusing to CD2 (CD2^{Ex}CD45^{Int}) blocks TCR triggering (ZAP70 remains cytosolic). Quantification is shown in **d**. Scale bars, 5 μ m.

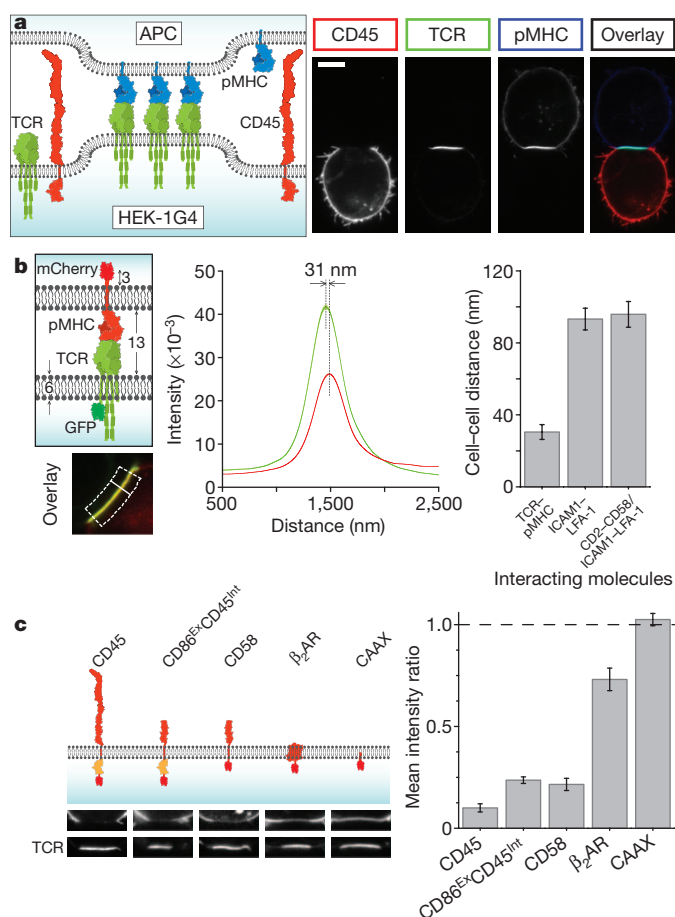


Figure 4 | The TCR-pMHC interaction drives protein exclusion at conjugate regions. **a**, A schematic and representative image data set showing that the TCR-pMHC interaction is sufficient to drive CD45 exclusion and its own clustering. Scale bar, 5 μ m. **b**, The inter-membrane distance between the conjugates (see Supplementary Methods) was measured, shown schematically as the separation between the two fluorophores over a normal line (white line) averaged across the conjugate region (dotted box). This procedure was performed for the cognate TCR-pMHC interaction ($n = 20$ cells), LFA-1-ICAM1 ($n = 23$ cells) and CD2-CD58 + LFA-1-ICAM1; $n = 20$ cells) interactions in the presence of control pMHC. Data represent mean \pm s.e.m. **c**, HEK cells were transfected with TCR-GFP and indicated molecule (fused to mCherry) and conjugated with APCs (CD45 phosphatase domains shown in orange). Representative images of the conjugate region are shown, with quantification of the ratio of fluorescence inside and outside of the interface. Data are mean \pm s.e.m. ($n = 20$) for each construct.

is necessary and sufficient for CD45 exclusion, with no requirement for downstream TCR triggering/signalling.

Next, we examined how the TCR-pMHC interaction affected the spacing of the two plasma membranes of the interacting cells. Using a subdiffraction-resolution method, we measured the separation between a GFP-tagged TCR in the HEK cells and mCherry-pMHC in the APC (Fig. 4b and Supplementary Methods). The measured distance of ~ 31 nm (Fig. 4b) suggests a cell-cell separation of ~ 15 nm, which agrees well with the 13 nm cell-to-cell distance between T cells and conjugated APCs measured by electron microscopy (ref. 28). By contrast, the membrane separation for Raji cells expressing control pMHC conjugated through adhesion molecules alone (ICAM1-LFA-1 or ICAM1-LFA-1 and CD2-CD58) was much greater (Fig. 4b). These results show that the TCR-pMHC interaction brings the membranes much closer together than occurs with the adhesion molecules.

Proteins with extended extracellular domains (such as CD45) might be prevented from entering regions of close membrane

apposition, as suggested by the kinetic-segregation model⁸. We tested this hypothesis by fusing the intracellular phosphatase domains of CD45 to an extracellular domain (from CD86) of comparable size to the TCR (termed CD86^{Ex}CD45^{Int}). CD86^{Ex}CD45^{Int} was also excluded from the cell-cell interface, although its exclusion (Fig. 4c) and TCR triggering (Supplementary Fig. 7a, b) were somewhat lower than that seen with the large CD45 construct (Fig. 4c). Because of this unanticipated exclusion of a protein with a small extracellular domain, we next tested mCherry fusion proteins of a series of membrane proteins with different properties: CD58, a transmembrane protein with an equivalently sized extracellular domain to CD86; β_2 -adrenergic receptor (β_2 AR, also known as ADRB2), a seven-transmembrane protein with small extracellular loops; and a prenylated version of the fluorophore linked to the inner leaflet of the bilayer by using a short targeting sequence (CAAX) (Fig. 4c). CD58 was excluded to a similar level as CD86^{Ex}CD45^{Int}, showing that the segregation of the latter construct was not due to the intracellular phosphatase domains of CD45 (Fig. 4c). β_2 AR, a protein with a large lateral footprint, was also partially excluded (Fig. 4c), possibly through a crowding effect arising from the high density of pMHC-TCR in the conjugate region. However, the prenylated mCherry was distributed evenly throughout the cell membrane, showing no exclusion by TCR-pMHC (Fig. 4c).

The above experiments cannot explain why CD2 (Supplementary Fig. 5b) and CD2^{Ex}CD45^{Int} (Fig. 2e), which has extracellular domains similar in size to CD86, were not excluded from the cell-cell contact zone. We speculated that binding of CD2 to its ligand (CD58) expressed endogenously on the APC provided a counteracting force to constrain CD2 within the conjugate region. If this were true, then expressing CD28, the binding partner for CD86, on the APC should diminish the TCR-pMHC-mediated segregation of CD86^{Ex}CD45^{Int}. Indeed, when CD86^{Ex}CD45^{Int}-expressing HEK-1G4 cells were conjugated with APCs co-expressing CD28, CD86^{Ex}CD45^{Int} localized at the cell-cell interface and TCR triggering was greatly decreased (Supplementary Fig. 7b, c).

In summary, the TCR-pMHC interaction alone is capable of excluding plasma membrane proteins with extracellular extensions and that exclusion can be overcome by the energy provided by binding to a protein partner on the APC.

Triggering of artificial receptors

The preceding experiments suggested that TCR triggering results from CD45 segregation, which is driven by the binding interaction between the TCR and pMHC. If this is true, then interactions between extracellular domains of membrane proteins with the proper spacing and affinity might elicit comparable effects to TCR-pMHC, as has been shown in T cells²⁸. To explore this idea, we engineered a chemically controlled, cell-surface-receptor system consisting of a transmembrane protein with extracellular FKBP and the intracellular CD3 ζ ITAM domains expressed in the HEK cell (FKBP^{Ex} ζ ^{Int}, mimicking the TCR) and a transmembrane protein with an extracellular FRB expressed in the APC (mimicking the pMHC) (Fig. 5a). FKBP^{Ex} ζ ^{Int} and FRB^{Ex} will only interact in the presence of rapamycin, forming a complex that spans a similar distance to TCR-pMHC. In the absence of rapamycin, conjugates formed (through the LFA-1-ICAM1 interaction) but ZAP70 was not recruited to the membrane (Fig. 5b, c). However, with rapamycin, FKBP^{Ex} ζ ^{Int} receptor clusters were observed at the cell-cell interface, even with low levels of FRB^{Ex} ligand that are equivalent to physiological densities of antigen pMHC (~ 5 molecules per μ m²; Supplementary Fig. 1). Furthermore, CD45 was excluded from and ZAP70 was recruited to these receptor clusters, indicating that triggering had occurred (Fig. 5b-d).

Our artificial receptor is structurally analogous to chimaeric antigen receptors (CARs), which have an extracellular single-chain antibody fragment fused to a transmembrane sequence and cytoplasmic CD3 ζ ITAMs. When the CD19-specific CAR (ref. 29), which has

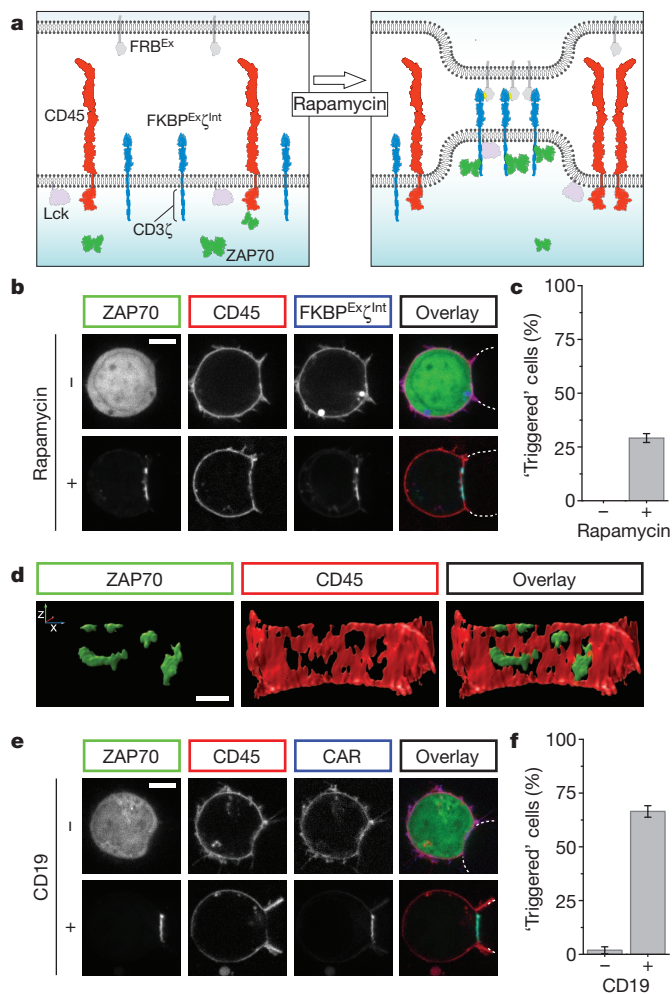


Figure 5 | Artificial receptor systems can cause CD45 exclusion and triggering. **a**, Schematic of the chemically inducible receptor system. FRB^{Ex} replaces pMHC on the APC and FKBP^{Ex/Int} replaces the TCR. Rapamycin induces FKBP^{Ex/Int}–FRB^{Ex} interaction. Additional molecules have been omitted for clarity. **b**, Rapamycin addition causes ZAP70 accumulation, denoting receptor triggering (gamma correction applied to lower images). FRB^{Ex}-expressing cells are shown by dotted lines. Scale bar, 5 μm. **c**, Quantification of rapamycin-induced triggering. Data are mean ± s.e.m. over four experiments. **d**, A deconvolved three-dimensional rendering of the reconstituted cell interface shown in **b**. Scale bar, 2 μm. **e**, Reconstituted HEK cells with the TCR replaced by a CAR specific for CD19 (see text) were conjugated with either CD19[−] (Jurkat) or CD19⁺ (Raji) cells, marked by dotted lines. Scale bar, 5 μm. **f**, Quantification of CAR-mediated triggering. Data are mean ± s.e.m. over three experiments.

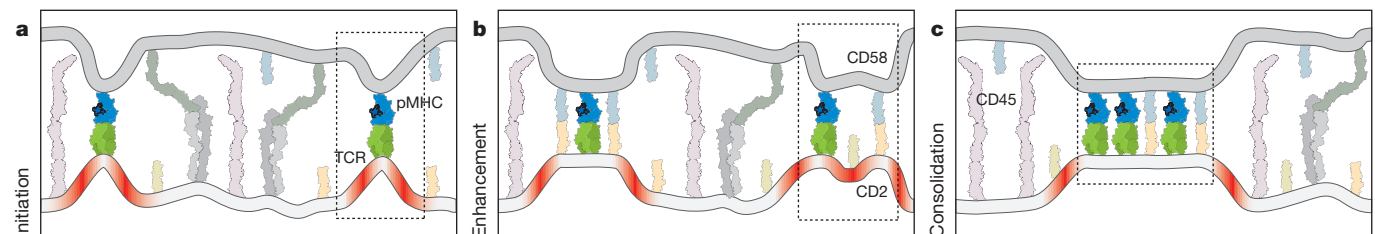


Figure 6 | A model for steps in TCR-mediated segregation based on membrane bending and energy minimization. The schematic uses equivalent molecule representations to previous figures (only extracellular domains are shown for simplicity) and boxed regions highlight features of each panel. **a**, After initial adhesion driven by large receptors such as LFA-1, transient fluctuations in the inter-membrane distance permit encounters between TCR and pMHC; the binding interaction overcomes energetically

shown promise in recent clinical trials as a gene therapy treatment of leukaemia²⁹, was expressed in our reconstituted HEK cells and these cells were conjugated with Raji B cells (which are CD19⁺), we observed CAR clustering, ZAP70 recruitment and CD45 exclusion (Fig. 5e, f). This triggering was ligand specific, as conjugation with CD19[−] cells did not produce these effects (Fig. 5e, f). Interestingly, the interface of many CD19–CAR HEK cell and CD19⁺ Raji B-cell conjugates showed a highly convoluted membrane surface (Supplementary Movie 5), which was not seen for cells interacting through either TCR–pMHC or FKBP^{Ex/Int}–FRB^{Ex}. This membrane effect could be a product of high-affinity antibody binding and might warrant further investigation for its relevance to CAR potency or potential side effects.

Conclusion

Our reconstitution experiments show a physical mechanism for TCR triggering that differs from dimerization or conformational-change models proposed for many cell-surface receptors. We find that the binding energy of the TCR–pMHC interaction generates an exclusion force for membrane proteins with large and/or unligated extracellular domains, even in the absence of downstream signalling. By linking an inhibitory phosphatase activity to a transmembrane protein (CD45) that is subject to the exclusion force and an activating kinase (Lck) to the inner leaflet of the membrane that is not, the TCR–pMHC interaction can shift the kinase–phosphatase balance and thus trigger the TCR, as first suggested by the kinetic-segregation model⁸. Because the FKBP–rapamycin–FRB or antibody–antigen modules can replace the extracellular TCR–pMHC interaction, conformational changes in the TCR are unlikely to be critical for the fundamental mechanism of triggering. Co-receptors and actin also do not seem to be essential for transducing TCR–pMHC binding across the plasma membrane, although it is very probable that they are necessary for achieving the high sensitivity and full selectivity of T-cell activation. This could be tested in the future by pushing our reconstituted system to respond to very low antigen densities using more sensitive read-outs of triggering.

The precise mechanism by which the TCR–pMHC interaction results in protein exclusion remains elusive, although our data provide certain clues. Exclusion according to the kinetic-segregation model is based on the greater size of the extracellular domain of CD45 compared to the TCR–pMHC complex. However, although size influences the extent of exclusion, our data show that membrane proteins with small extracellular domains are also excluded, suggesting that additional forces must be acting on the system. We speculate upon the driving force in the following model, which incorporates previous findings^{30–34}. Within the initial adhesion mediated by LFA-1–ICAM1 between the T cell and APC, transient fluctuations bring the two membranes in closer apposition, allowing TCR and pMHC molecules to interact (Fig. 6a). The binding energy of this interaction must be sufficient to overcome unfavourable membrane bending and compression of the large proteins that constitute the

unfavourable membrane bending (red regions). **b**, Additional molecules (such as CD2 and CD58) provide additional binding energy that stabilizes regions of local membrane bending and may enhance the local exclusion of larger molecules by TCR–pMHC interactions. **c**, Consolidation of discrete contact regions serves to minimize unfavourable membrane bending and leads to the exclusion of smaller proteins that do not provide any countering ligand-binding free energy. See text for details.

glycocalyx, such as CD45, which occur when the two membranes regions are brought close together (Fig. 6b). In a critical next step, separate regions of close membrane apposition consolidate passively into larger, contiguous regions, resulting in a net decrease in membrane bending, as has been previously suggested³⁵. Proteins that do not provide any binding energy will be excluded over time, because, as they diffuse from the region, interacting molecules will be further clustered as the area of close apposition is minimized (Fig. 6c). These same principles may explain how TCR microdomains initiate when T cells interact with pMHC on a supported lipid bilayer^{36–38} (note the small exclusion zones in our experiments at low ligand densities (Fig. 5d and Supplementary Fig. 4)).

Overall, our data suggest how the binding energy associated with specific molecular recognition events that take place between two interacting cells can be transduced into intracellular biochemical reactions that change cell behaviour. This model provides a plausible mechanism to explain how CARs trigger T cells to kill cancerous cells²⁹ and may apply to other cell types that signal using membrane-bound receptors and ligands.

METHODS SUMMARY

Multiple proteins were expressed in HEK cells using a combination of transient and stable expression by lentiviral transduction; proteins were expressed at close to physiological levels and in the correct localization as described in Supplementary Methods. To create cell conjugates, the two cell types were centrifuged, resuspended at high density and placed on glass-bottomed dishes for live-cell imaging at 37 °C using spinning disc confocal microscopy. Image analysis, including the inter-membrane distance algorithm, was performed using ImageJ and Matlab (Supplementary Methods). Full methods and supplementary material accompany this paper.

Received 14 November 2011; accepted 8 May 2012.

Published online 27 June 2012.

- Smith-Garvin, J. E., Koretzky, G. A. & Jordan, M. S. T cell activation. *Annu. Rev. Immunol.* **27**, 591–619 (2009).
- Lemmon, M. A. & Schlessinger, J. Cell signaling by receptor tyrosine kinases. *Cell* **141**, 1117–1134 (2010).
- Palacios, E. H. & Weiss, A. Function of the Src-family kinases, Lck and Fyn, in T-cell development and activation. *Oncogene* **23**, 7990–8000 (2004).
- Love, P. E. & Hayes, S. M. ITAM-mediated signaling by the T-cell antigen receptor. *Cold Spring Harb. Perspect. Biol.* **2**, a002485 (2010).
- Au-Yeung, B. B. *et al.* The structure, regulation, and function of ZAP-70. *Immunol. Rev.* **228**, 41–57 (2009).
- van der Merwe, P. A. & Dushek, O. Mechanisms for T cell receptor triggering. *Nature Rev. Immunol.* **11**, 47–55 (2011).
- Xu, C. *et al.* Regulation of T cell receptor activation by dynamic membrane binding of the CD3 ϵ cytoplasmic tyrosine-based motif. *Cell* **135**, 702–713 (2008).
- Davis, S. J. & van der Merwe, P. A. The kinetic-segregation model: TCR triggering and beyond. *Nature Immunol.* **7**, 803–809 (2006).
- Lillemeier, B. F. *et al.* TCR and Lat are expressed on separate protein islands on T cell membranes and concatenate during activation. *Nature Immunol.* **11**, 90–96 (2010).
- Varma, R., Campi, G., Yokosuka, T., Saito, T. & Dustin, M. L. T cell receptor-proximal signals are sustained in peripheral microclusters and terminated in the central supramolecular activation cluster. *Immunity* **25**, 117–127 (2006).
- Szymczak, A. L. *et al.* Correction of multi-gene deficiency *in vivo* using a single ‘self-cleaving’ 2A peptide-based retroviral vector. *Nature Biotechnol.* **22**, 589–594 (2004).
- Aleksic, M. *et al.* Dependence of T cell antigen recognition on T cell receptor–peptide MHC confinement time. *Immunity* **32**, 163–174 (2010).
- Holst, J. *et al.* Scalable signaling mediated by T cell antigen receptor-CD3 ITAMs ensures effective negative selection and prevents autoimmunity. *Nature Immunol.* **9**, 658–666 (2008).
- Irving, B. A. & Weiss, A. The cytoplasmic domain of the T cell receptor ζ chain is sufficient to couple to receptor-associated signal transduction pathways. *Cell* **64**, 891–901 (1991).
- Deindl, S. *et al.* Structural basis for the inhibition of tyrosine kinase activity of ZAP-70. *Cell* **129**, 735–746 (2007).
- Bergman, M. *et al.* The human p50csk tyrosine kinase phosphorylates p56lck at Tyr-505 and down regulates its catalytic activity. *EMBO J.* **11**, 2919–2924 (1992).
- Hermiston, M. L., Xu, Z. & Weiss, A. CD45: a critical regulator of signaling thresholds in immune cells. *Annu. Rev. Immunol.* **21**, 107–137 (2003).
- Saunders, A. E. & Johnson, P. Modulation of immune cell signalling by the leukocyte common tyrosine phosphatase, CD45. *Cell. Signal.* **22**, 339–348 (2010).
- Chen, J. L. *et al.* Structural and kinetic basis for heightened immunogenicity of T cell vaccines. *J. Exp. Med.* **201**, 1243–1255 (2005).
- Monks, C. R., Freiberg, B. A., Kupfer, H., Sciaky, N. & Kupfer, A. Three-dimensional segregation of supramolecular activation clusters in T cells. *Nature* **395**, 82–86 (1998).
- Altan-Bonnet, G. & Germain, R. N. Modeling T cell antigen discrimination based on feedback control of digital ERK responses. *PLoS Biol.* **3**, e356 (2005).
- Manz, B. N., Jackson, B. L., Petit, R. S., Dustin, M. L. & Groves, J. T-cell triggering thresholds are modulated by the number of antigen within individual T-cell receptor clusters. *Proc. Natl Acad. Sci. USA* **108**, 9089–9094 (2011).
- Valitutti, S., Dessing, M., Aktories, K., Gallati, H. & Lanzavecchia, A. Sustained signaling leading to T cell activation results from prolonged T cell receptor occupancy. Role of T cell actin cytoskeleton. *J. Exp. Med.* **181**, 577–584 (1995).
- Johnson, K. G., Bromley, S. K., Dustin, M. L. & Thomas, M. L. A supramolecular basis for CD45 tyrosine phosphatase regulation in sustained T cell activation. *Proc. Natl Acad. Sci. USA* **97**, 10138–10143 (2000).
- Leupin, O., Zaru, R., Laroche, T., Muller, S. & Valitutti, S. Exclusion of CD45 from the T-cell receptor signaling area in antigen-stimulated T lymphocytes. *Curr. Biol.* **10**, 277–280 (2000).
- Irles, C. *et al.* CD45 ectodomain controls interaction with GEMs and Lck activity for optimal TCR signaling. *Nature Immunol.* **4**, 189–197 (2003).
- He, X., Woodford-Thomas, T. A., Johnson, K. G., Shah, D. D. & Thomas, M. L. Targeting of CD45 protein tyrosine phosphatase activity to lipid microdomains on the T cell surface inhibits TCR signaling. *Eur. J. Immunol.* **32**, 2578–2587 (2002).
- Choudhuri, K., Wiseman, D., Brown, M. H., Gould, K. & van der Merwe, P. A. T-cell receptor triggering is critically dependent on the dimensions of its peptide-MHC ligand. *Nature* **436**, 578–582 (2005).
- Porter, D. L., Levine, B. L., Kalos, M., Bagg, A. & June, C. H. Chimeric antigen receptor-modified T cells in chronic lymphoid leukemia. *N. Engl. J. Med.* **365**, 725–733 (2011).
- Shaw, A. S. & Dustin, M. L. Making the T cell receptor go the distance: a topological view of T cell activation. *Immunity* **6**, 361–369 (1997).
- Qi, S. Y., Groves, J. T. & Chakraborty, A. K. Synaptic pattern formation during cellular recognition. *Proc. Natl Acad. Sci. USA* **98**, 6548–6553 (2001).
- Weikl, T. R. & Lipowsky, R. Pattern formation during T-cell adhesion. *Biophys. J.* **87**, 3665–3678 (2004).
- Coombs, D., Dembo, M., Wofsy, C. & Goldstein, B. Equilibrium thermodynamics of cell-cell adhesion mediated by multiple ligand–receptor pairs. *Biophys. J.* **86**, 1408–1423 (2004).
- Alakoskela, J. M. *et al.* Mechanisms for size-dependent protein segregation at immune synapses assessed with molecular rulers. *Biophys. J.* **100**, 2865–2874 (2011).
- Burroughs, N. J. & Wulfig, C. Differential segregation in a cell–cell contact interface: the dynamics of the immunological synapse. *Biophys. J.* **83**, 1784–1796 (2002).
- Yokosuka, T. *et al.* Newly generated T cell receptor microclusters initiate and sustain T cell activation by recruitment of Zap70 and SLP-76. *Nature Immunol.* **6**, 1253–1262 (2005).
- Campi, G., Varma, R. & Dustin, M. L. Actin and agonist MHC–peptide complex-dependent T cell receptor microclusters as scaffolds for signaling. *J. Exp. Med.* **202**, 1031–1036 (2005).
- Bunnell, S. C. *et al.* T cell receptor ligation induces the formation of dynamically regulated signaling assemblies. *J. Cell Biol.* **158**, 1263–1275 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank A. van der Merwe and V. Cerundolo for the 1G4 TCR sequence, A. Weiss for cell lines and advice, C. June for the CD19 CAR construct, N. Stuurman and K. Thorn for microscopy help and members of the Vale laboratory for discussions. R.D.V. is a Howard Hughes Medical Institute investigator and J.R.J. is a fellow of the Jane Coffin Childs Memorial Fund.

Author Contributions J.R.J. conceived the study, collected the data and conducted the analyses. J.R.J. and R.D.V. designed the experiments and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to R.D.V. (vale@cmp.ucsf.edu).

Direct and highly regioselective and enantioselective allylation of β -diketones

Wesley A. Chalifoux¹, Samuel K. Reznik¹ & James L. Leighton¹

The enantioselective allylation of ketones is a problem of fundamental importance in asymmetric reaction design, especially given that only a very small number of methods can generate tertiary carbinols. Despite the vast amount of attention that synthetic chemists have given to this problem^{1–8}, success has generally been limited to just a few simple ketone types. A method for the selective allylation of functionally complex ketones would greatly increase the utility of ketone allylation methods in the chemical synthesis of important targets. Here we describe the operationally simple, direct, regioselective and enantioselective allylation of β -diketones. The strong tendency of β -diketones to act as nucleophilic species was overcome by using their enol form to provide the necessary Brønsted-acid activation. This reaction significantly expands the pool of enantiomerically enriched and functionally complex tertiary carbinols that may be easily accessed. It also overturns more than a century of received wisdom regarding the reactivity of β -diketones.

Chiral non-racemic tertiary carbinols are common in important natural products and in medicinal-chemistry research programmes. Asymmetric ketone allylation is one of the few methods available for their direct synthesis, and over the past 15 years there have been several seminal reports that describe important and significant progress towards the development of efficient, practical and highly enantioselective techniques^{1–8}. With only a few exceptions, however, these methods are limited in scope to aryl methyl ketones and aryl linear alkyl ketones (that is, ArCOCH_2R , where Ar is an aryl group and R is hydrogen, an alkyl group or an aryl group). Thus, although efficiency and practicality rightly constitute one major focus in asymmetric reaction design, scope, generality, reliability and applicability in complex settings are equally vital to the realization of methods with truly broad utility and that will be widely adopted⁹.

These thoughts were triggered by our retrosynthesis of the AB spiroketal portion of the potent and precious anti-mitotic marine macrolide spongistatin 1^{10–12}—which bears a tertiary carbinol—to β -hydroxyketone **1** (Fig. 1a). It seemed that the ideal¹³ synthesis of **1** should involve a direct asymmetric allylation of acetylacetone (Fig. 1b). However, although a few sporadic examples of the non-asymmetric allylation of acetylacetone have been reported^{14–16}, there are no known examples of an enantioselective addition of a nucleophile to a β -diketone. In fact, there are simply no general methods that use β -diketones as electrophiles in reactions to form carbon–carbon bonds. This is unsurprising given that β -diketones are not simply compounds with two ketones, but rather are distinct chemical entities that exist primarily in their enol tautomer form. They are rarely used in organic chemical synthesis, and when they are, it is typically as nucleophiles in reactions such as simple alkylations or Knoevenagel condensations. As a result, the synthesis of **1** by means of an asymmetric allylation approach using known methods would require protection or masking of one of the ketones, thus turning what is in principle an ideal synthesis into a multi-step process that would in practice involve protecting-group manipulations and/or oxidation-state adjustments. The development of a method for the direct asymmetric allylation of

acetylacetone seemed a worthy enough aim, but the ultimate goal of this study became the generalization of such a method to a broad range of β -diketones, including unsymmetrical β -diketones. Far from representing a straightforward extension of the proposed method, however, the allylation of unsymmetrical β -diketones poses a further distinct and formidable challenge: the control of regioselectivity (Fig. 1c). Intrigued by these challenges and by the almost total dearth of useful precedent, we have pursued a general, practical and highly regioselective and enantioselective allylation of β -diketones. We report its successful development and mechanistic elucidation.

Rather than attempting to mask the enol tautomer of the β -diketone *in situ*, preclude its formation or otherwise render it innocuous, our approach centred on the idea that we might instead react it with an allylchlorosilane to generate β -silyloxyenone **2**, which would subsequently undergo intramolecular allylation (Fig. 2a). However, this otherwise beneficial tethering strategy would in this case also lead to deactivation of the ketone electrophile by converting it into a vinylogous silyl ester, raising concerns about whether the desired reaction might be impractically slow and allow competing side reactions (for example, Knoevenagel condensation between **2** and unreacted diketone, or intermolecular diketone allylation by the allylsilane or by **2**). Ideally, then, to favour the desired pathway, the complexation should be very fast, so that no unreacted diketone is available for such side reactions, and the allylsilane should become strongly activated upon complexation. Allylsilane **3** and other related aminosilanes are strongly activated by protonation with Brønsted acids^{17,18}. Given that the proposed enol silylation reaction would produce an equivalent of hydrochloric acid, it seemed possible that

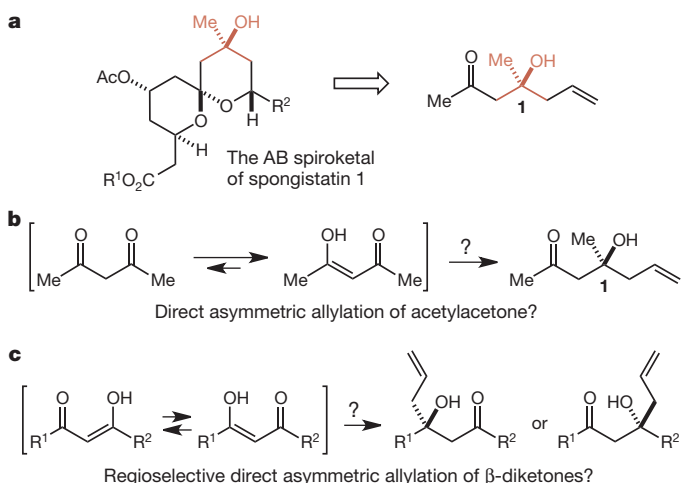


Figure 1 | Origin and evolution of a proposal for a direct, enantioselective and regioselective allylation of β -diketones. **a**, Retrosynthesis of the AB spiroketal of spongistatin **1** to β -hydroxyketone **1**. Me, methyl group; Ac, acetyl group; R, unspecified group. **b**, A direct asymmetric allylation of acetylacetone would have to contend with the enol form of the diketone. **c**, The generalization of a direct method would require control of regioselectivity.

¹Department of Chemistry, Columbia University, New York, New York 10027, USA.

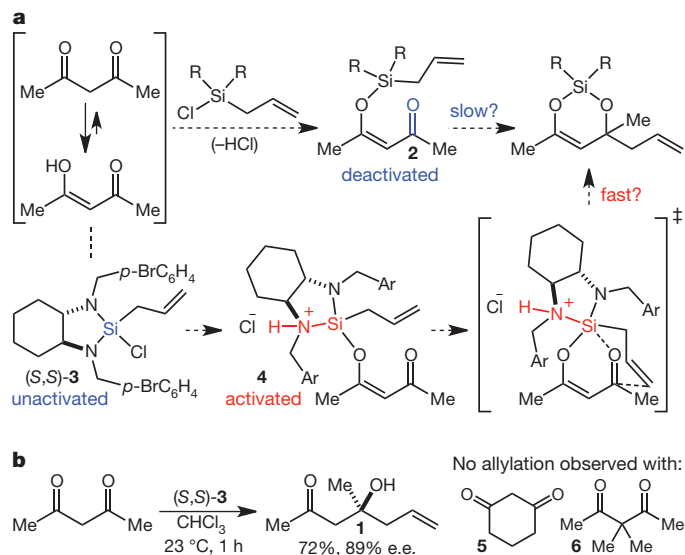
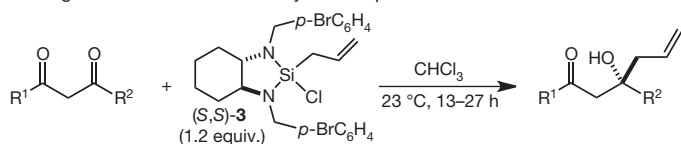


Figure 2 | Reaction design and proof of concept. **a**, The reaction design entails tethering the allylsilane to the enol form of the diketone and activation of the silane by the liberated hydrochloric acid (HCl). Ar, aryl group. **b**, First demonstration of the asymmetric allylation of acetylacetone.

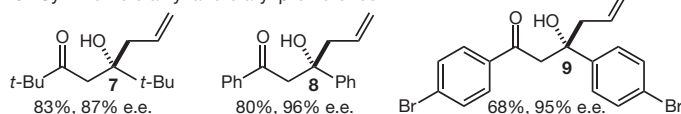
3, which is wholly incapable of allylating ketones, might be perfectly configured to react with β -diketones by way of activated silane complex **4** (Fig. 2a). Indeed, it transpired that allylsilane **3** smoothly allylated acetylacetone, and, on optimization of the reaction conditions (CHCl_3 , 23°C , 1 h), produced **1** in 72% yield and 89% enantiomeric excess (e.e.) (Fig. 2b). This operationally simple reaction provides direct and scalable access to **1** from two commercially available compounds, and neither **5** nor **6** was allylated by **3** (Supplementary Fig. 1 and 2), which strongly implies that the underlying mechanistic proposal is sound. The enol form of the diketone, which usually represents a serious functional-group incompatibility, has thus been co-opted as the very functionality responsible for the success of the reaction.

With this proof of concept in hand, we examined the scope of the reaction (Fig. 3a). We began with three further symmetrical β -diketones; as shown (Fig. 3b), products **7**, **8** and **9** were all obtained in good yields and with good to excellent enantioselectivity, establishing that both *t*-butyl and aryl ketones are well tolerated. We next examined a series of unsymmetrical β -diketones, in an effort to establish whether and under what circumstances useful levels of regioselectivity could be realized. Whereas substrates that pitted two alkyl ketones against each other led to little or no regioselectivity, we quickly discovered that substrates that pitted a conjugated ketone against an alkyl ketone consistently reacted such that the allyl group was added to the alkyl ketone with excellent regio- and enantioselectivity (Fig. 3c). This preference not only led to the efficient and selective production of **10–12**, but also operated even when the alkyl group is a *t*-butyl group (as in **13**). Next, we examined 1,3-diphenylpropane-1,3-diones with varying substitutions. We were disappointed at first to find that the *para*-bromo, *para*-methoxy derivative was completely non-selective. When we interrogated the impact of *ortho* substitution, however, subtle effects were revealed (Fig. 3d). Thus, whereas an *ortho*-bromo group regioselectively directs the allylation to the distal ketone (as in **14** and **15**), an *ortho*-methoxy group has the opposite effect (as in **16**) albeit with only moderate (4:1) regioselectivity. These effects are mutually reinforcing, as demonstrated by the reaction of the *ortho*-bromo, *ortho*-methoxy analogue to give **17** as a single regioisomer in 89% yield and 95% e.e., a result that is particularly striking in light of the complete lack of regioselectivity in the reaction of the *para*, *para* isomer. Two final examples (**18** and **19**) show that this *ortho*-effect extends to groups other than bromide, and establish that the methoxy group is the outlier (as in **16**).

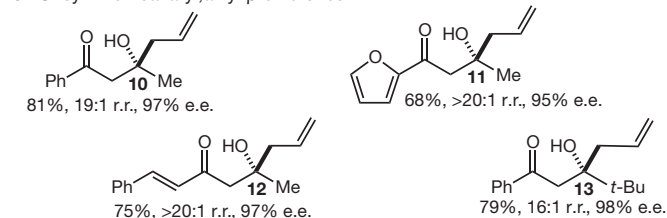
a Regio- and enantioselective allylation of β -diketones



b Symmetrical dialkyl and diaryl β -diketones



c Unsymmetrical aryl, alkyl β -diketones



d Unsymmetrical diaryl β -diketones

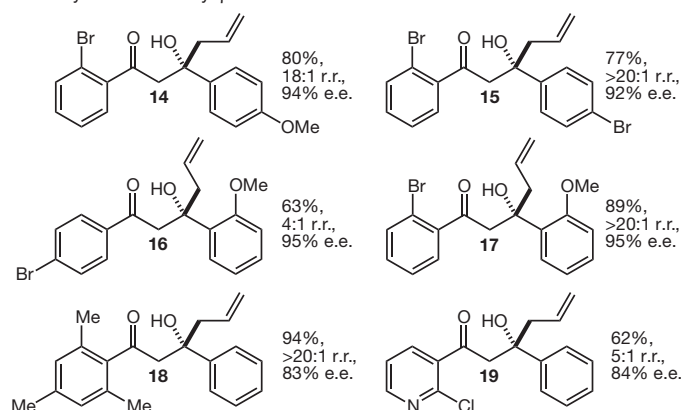


Figure 3 | Scope of the regio- and enantioselective allylation of β -diketones. **a**, The β -diketone was treated with 1.2 equiv. of (S,S)-**3** in chloroform (CHCl_3) and the resultant mixture was stirred at ambient temperature for 13–27 h. In all cases, the reported yield refers to the yield of isolated (greater than or equal to 20:1 regioisomeric purity) major product after workup and purification by flash chromatography, and the enantiomeric excess (e.e.) was determined by chiral high-performance liquid chromatography. d.r., diastereomer ratio. **b**, Enantioselective allylation of symmetrical β -diketones. Ph, phenyl group. **c**, Regio- and enantioselective allylation of unsymmetrical aryl, alkyl β -diketones (r.r., regioisomer ratio). **d**, Regio- and enantioselective allylation of unsymmetrical diaryl β -diketones.

In principle, the extension of this reaction methodology to crotylation reactions would allow the establishment of a second stereocentre in the allylic position of the homoallylic alcohol products, but examples of highly diastereo- and enantioselective ketone crotylation reactions are rare^{5,7,8,19}. We therefore decided to interrogate the ability of crotylsilanes **20** and **21** (ref. 20) to crotylate β -diketones, and found that treatment of acetylacetone and benzoylacetone with *trans*-crotylsilane **20** led to the isolation of **22** (69%, 89% e.e.) and **23** (75%, 97% e.e.) respectively (Fig. 4a). However, treatment of benzoylacetone with *cis*-crotylsilane **21** resulted in a much slower reaction that provided the desired product in very low yield (less than 15%), along with significant amounts of a by-product that we have tentatively identified as **24**. The formation of **24** is readily rationalized as the result of a Knoevenagel condensation cascade that presumably begins with attack of the silyl enol ether complex on unreacted β -diketone. This hypothesis implied that the complexation event is slow in these reactions in general, and that the desired carbon–carbon

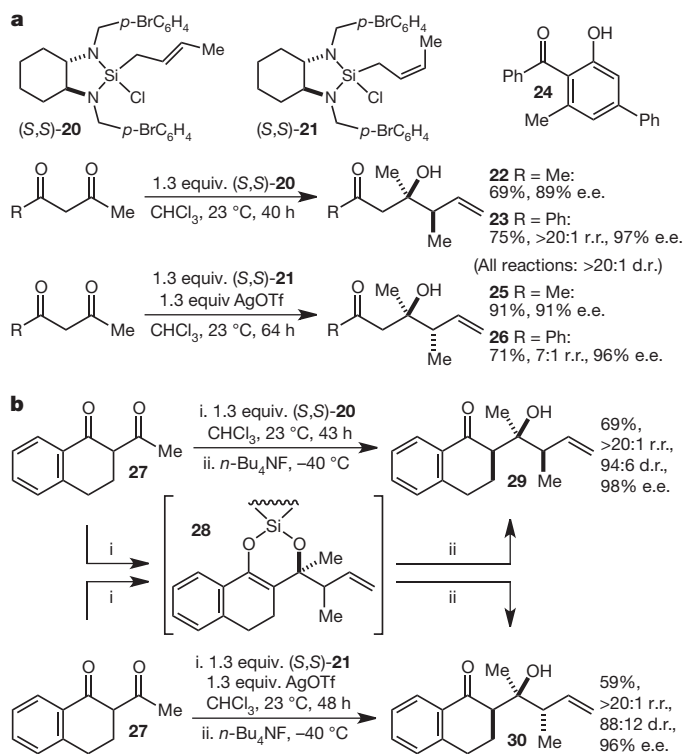


Figure 4 | **Regio-, diastereo- and enantioselective crotylation of β -diketones.** **a**, The crotylation of β -diketones allows the establishment of a second stereocentre in the homoallylic alcohol products. OTf, triflate. **b**, α -Substitution on the β -diketone substrate may be parlayed into reactions that establish three stereocentres from simple starting materials.

bond-forming step is especially slow in this particular reaction. In this way, the silyl enol ether complex and the starting diketone would both be present in significant concentrations for long periods of time. One way to avoid this could be to accelerate the complexation so that no unreacted diketone would be available to the silyl enol ether complex, and it seemed plausible that replacement of the chloride on the silane with a triflate ($^-OSO_2CF_3$) might accomplish this. Indeed, pretreatment of **21** with $AgOSO_2CF_3$ before the addition of acetylacetone and benzoylacetone led to vastly improved results, giving **25** and **26** respectively, in excellent yields and enantioselectivities.

Having worked out effective conditions for the crotylation reactions, we next examined 2-acetyltetralone **27** to probe the effect of substitution on the α -carbon (Fig. 4b). It will be noted that the initial products in these reactions are silyl enol ethers (such as **28**), which are converted to enols that tautomerize to ketones on quench and workup. When the α -carbon is substituted, as in **28**, the tautomerization has stereochemical consequences, and we have demonstrated in similar systems that good to excellent diastereocontrol may be realized with proper choice of quench/workup conditions²¹. The standard quench for the present method entails cooling the reaction mixtures to $-40^\circ C$ and adding $n-Bu_4NF$ (where Bu is a butyl group), followed by an aqueous workup. When we quenched and worked up the crotylation reactions of **27** with **20** and **21** using these conditions, good to excellent diastereoselectivity was realized, resulting in the isolation of **29** and **30** in 69% and 59% yield, respectively, with excellent regio- and enantioselectivity. The result is an operationally simple single-step and protecting-group-free synthesis of complex polyketide-like arrays of three stereocentres (including a tertiary carbinol) from simple and readily available starting materials with exceedingly high levels of enantioselectivity.

We next tried to elucidate the mechanistic basis for the remarkable regioselectivity of the reaction. The reaction of an unsymmetrical β -diketone with (*S,S*)-**3** (Fig. 5a) may produce four different

β -silyloxyenones (**31E**, **31Z**, **32E** and **32Z**), which may be considered as two regioisomeric pairs of *E* and *Z* isomers. Once it has been established that the *E* and *Z* isomers in each pair interconvert quickly (Supplementary Fig. 3), there are two readily apparent possibilities. The first is that there is no (fast) interconversion between **31Z** and **32Z**, and the regioselectivity is determined by the relative rates of formation of **31** and **32**. The second is that interconversion of **31Z** and **32Z**, presumably by way of **33**, is fast and the regioselectivity is determined by the relative energies of transition states **34** and **35**. To distinguish between these possibilities, we repeated the reaction of **27** with **21** (Fig. 4b) in deuterated chloroform, and monitored it by 1H NMR spectroscopy (Supplementary Fig. 4). After one hour, the singlet for the methyl group in **27** disappeared and three new singlets appeared at 2.53, 2.11 and 1.97 p.p.m. in a roughly 2:3:4:1 ratio. We have assigned these peaks to the three possible silyl enol ether complexes. Given that the peaks all disappear at the same rate and that this reaction is completely regioselective despite only one of these silyl enol ethers being

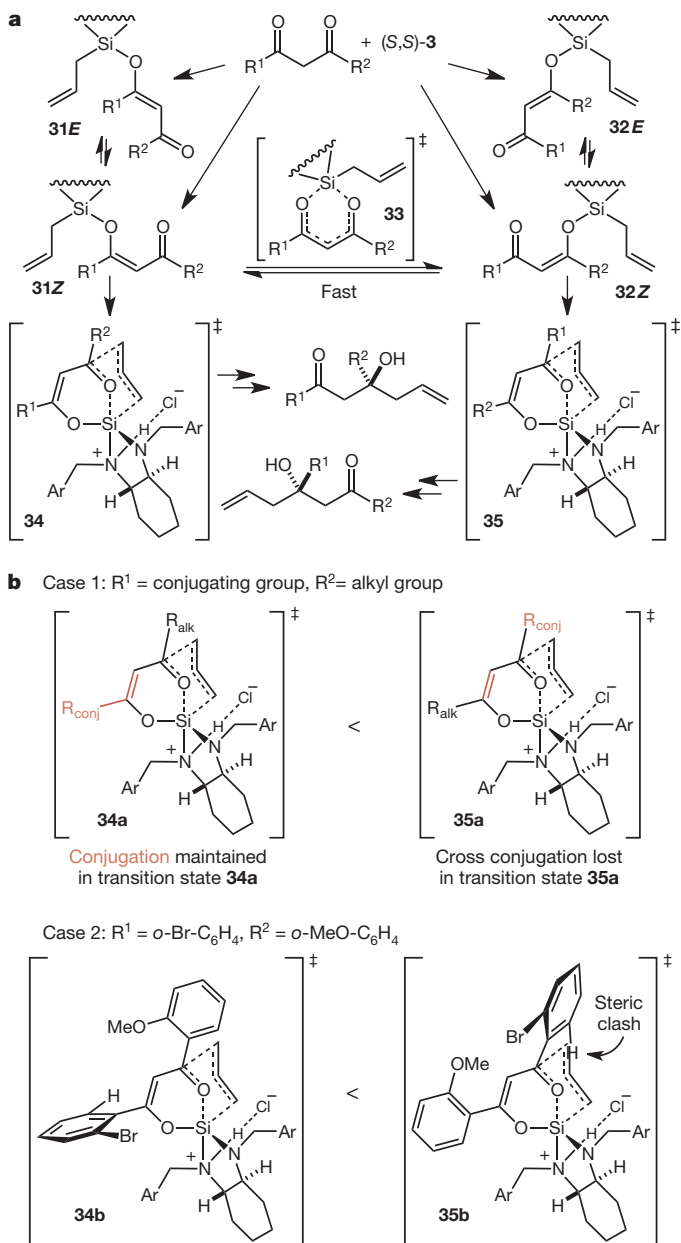


Figure 5 | **General mechanism and origin of regioselectivity.** **a**, A general mechanistic scheme for the allylation of β -diketones wherein the product distribution is determined by Curtin–Hammett kinetics. **b**, Models for the observed regioselectivity.

able to react directly to give the product, we conclude that in these reactions, the interconversion between **31Z** and **32Z** is fast^{22–25} and the regioselectivity is determined by the relative energies of transition states **34** and **35**; that is, the reaction is governed by Curtin–Hammett kinetics²⁶.

The regioselectivity observed with substrates wherein $R^1 = \text{aryl}$ /alkenyl and $R^2 = \text{alkyl}$ (as in Fig. 3c) is difficult to attribute to a simple steric effect, given the regioselective production of **13**. Rather, we propose that the energy of transition state **34a** is lower than that for **35a** because conjugation is maintained in the former pathway but lost completely in the latter (Fig. 5b, case 1). Conversely, we propose that the regioselectivity observed with substrates wherein both R^1 and R^2 are aryl groups and at least one bears *ortho*-substitution (as in Fig. 3d) is the result of steric effects (Fig. 5b, case 2). We expect 2-bromophenyl, mesityl and 2-chloro-3-pyridyl groups to be rotated substantially out of conjugation with the ketones to which they are attached^{27–29}, rendering them highly sterically hindered towards nucleophilic attack, as shown in transition state **35b**. By contrast, 2-methoxybenzoyl and *ortho*-unsubstituted phenyl groups are conjugated³⁰, allowing an unhindered approach of the allyl group as in transition state **34b**. Thus, in this context, an *ortho*-bromophenyl group is effectively large and an *ortho*-methoxyphenyl group is effectively small. We believe that this steric effect is large enough to overwhelm the complete loss of conjugation in transition state **34b**. The moderate regioselectivity for product **16** is more difficult to rationalize, but it may be that the activation energies for the competing carbon–carbon bond-forming steps are nearly equivalent and the selectivity derives from a modest shifting of the equilibrium constant *K* towards **31Z** owing to the silane's preference for residing away from the larger *ortho*-methoxyphenyl group.

We have demonstrated that it is possible to achieve not only the highly enantioselective allylation and crotylation of β -diketones, but also the highly regioselective allylation and crotylation of unsymmetrical β -diketones. The method allows the protecting-group-free, single-step (that is, ideal) synthesis of functionally and stereochemically complex products from readily available—and undifferentiated— β -diketone starting materials. We have elucidated important aspects of the mechanism and found that the regioselectivity is governed by Curtin–Hammett kinetics. This may have important implications for the extension of this methodology to other dicarbonyl substrate types; experiments along these lines are in progress.

METHODS SUMMARY

The general procedure for the allylation reactions described in Fig. 3 is as follows. To a 0.10 M solution of β -diketone (1.0 equiv.) in anhydrous CHCl_3 is added (*S,S*)-**3** (1.2 equiv.). The resulting mixture is stirred at ambient temperature (roughly 23 °C) for between 12 and 27 h. The mixture is cooled to –40 °C, *n*-Bu₄NF (4.0 equiv., 1 M in tetrahydrofuran) is added, and the resulting mixture is maintained at –40 °C for 1 h. Saturated aqueous NH_4Cl is added and the mixture is allowed to warm to ambient temperature. The mixture is extracted three times with CH_2Cl_2 . The combined organic layers are washed with water and brine, dried over MgSO_4 , filtered and concentrated. The residue is purified by flash chromatography on silica gel. For complete experimental details and compound characterization, see Supplementary Information.

Received 5 April; accepted 4 May 2012.

Published online 27 June 2012.

1. Nakamura, M., Hirai, A., Sogi, M. & Nakamura, E. Enantioselective addition of allylzinc reagent to alkynyl ketones. *J. Am. Chem. Soc.* **120**, 5846–5847 (1998).
2. Waltz, K. M., Gavenonis, J. & Walsh, P. J. A simple, reliable, catalytic asymmetric allylation of ketones. *Angew. Chem. Int. Edn* **41**, 3697–3699 (2002).
3. Wu, T. R., Shen, L. & Chong, J. M. Asymmetric allylboration of aldehydes and ketones using 3,3'-disubstitutedbinaphthol-modified boronates. *Org. Lett.* **6**, 2701–2704 (2004).
4. Canales, E., Prasad, K. G. & Soderquist, J. A. *B*-Allyl-10-Ph-9-borabicyclo[3.3.2]decane: strategically designed for the asymmetric allylboration of ketones. *J. Am. Chem. Soc.* **127**, 11572–11573 (2005).
5. Wadamoto, M. & Yamamoto, H. Silver-catalyzed asymmetric Sakurai–Hosomi allylation of ketones. *J. Am. Chem. Soc.* **127**, 14556–14557 (2005).
6. Miller, J. J. & Sigman, M. S. Design and synthesis of modular oxazoline ligands for the enantioselective chromium-catalyzed addition of allyl bromide to ketones. *J. Am. Chem. Soc.* **129**, 2752–2753 (2007).

7. Barnett, D. S., Moquist, P. N. & Schaus, S. E. The mechanism and an improved asymmetric allylboration of ketones catalyzed by chiral biphenols. *Angew. Chem. Int. Edn* **48**, 8679–8682 (2009).
8. Shi, S.-L., Xu, L.-W., Oisaki, K., Kanai, M. & Shibasaki, M. Identification of modular chiral bisphosphines effective for Cu(I)-catalyzed asymmetric allylation and propargylation of ketones. *J. Am. Chem. Soc.* **132**, 6638–6639 (2010).
9. Robak, M. T., Herbage, M. A. & Ellman, J. A. Synthesis and applications of *tert*-butanesulfonimide. *Chem. Rev.* **110**, 3600–3740 (2010).
10. Pettit, G. R. *et al.* Isolation and structure of spongistatin 1. *J. Org. Chem.* **58**, 1302–1304 (1993).
11. Fusetani, N., Shinoda, K. & Matsunaga, S. Cinachryolide A: a potent cytotoxic macrolide possessing two spiro ketals from marine sponge *Cinachyra* sp. *J. Am. Chem. Soc.* **115**, 3977–3981 (1993).
12. Kobayashi, M. *et al.* Althohyrtin A, a potent anti-tumor macrolide from the Okinawan marine sponge *Hyrtios altum*. *Tetrahedr. Lett.* **34**, 2795–2798 (1993).
13. Gaich, T. & Baran, P. S. Aiming for the ideal synthesis. *J. Org. Chem.* **75**, 4657–4673 (2010).
14. Deng, D.-L. & Lu, Z.-H. Monoallylation of symmetric diketones in the presence of water by allylic bromide and metallic zinc. *Chin. Chem. Lett.* **5**, 173–176 (1994).
15. Kira, M., Sato, K., Sekimoto, K., Gewald, R. & Sakurai, H. Stereoselective allylation of β -hydroxy- and β -amino- α,β -enones with allyltrifluorosilane/triethylamine systems. *Chem. Lett.* **1995**, 281–282 (1995).
16. Marton, D., Stivanello, D. & Tagliavini, G. Allylstannation of α -, β - and γ -diketones mediated by allylbutyltin halides: $\text{Bu}_2(\text{CH}_2=\text{CHCH}_2)\text{SnCl}$ and $\text{Bu}(\text{CH}_2=\text{CHCH}_2)\text{SnCl}_2$. *J. Organomet. Chem.* **540**, 77–81 (1997).
17. Leighton, J. L. Powerful and versatile silicon Lewis acids for asymmetric chemical synthesis. *Aldrichim. Acta* **43**, 3–12 (2010).
18. Kim, H., Ho, S. & Leighton, J. L. A more comprehensive and highly practical solution to enantioselective aldehyde crotylation. *J. Am. Chem. Soc.* **133**, 6517–6520 (2011).
19. Burns, N. Z., Hackman, B. M., Ng, P. Y., Powelson, I. A. & Leighton, J. L. The enantioselective allylation and crotylation of sterically hindered and functionalized aryl ketones: convenient access to unusual tertiary carbinol structures. *Angew. Chem. Int. Edn* **45**, 3811–3813 (2006).
20. Hackman, B. M., Lombardi, P. J. & Leighton, J. L. Highly diastereo- and enantioselective reagents for aldehyde crotylation. *Org. Lett.* **6**, 4375–4377 (2004).
21. Spletstoser, J. T., Zacuto, M. J. & Leighton, J. L. Tandem silylformylation–crotylsilylation/Tamao oxidation of internal alkynes: a remarkable example of generating complexity from simplicity. *Org. Lett.* **10**, 5593–5596 (2008).
22. Pinnavaia, T. J., Collins, W. T. & Howe, J. G. Triorganosilicon acetylacetonates. Enol ether isomerism and stereochemical lability. *J. Am. Chem. Soc.* **92**, 4544–4550 (1970).
23. Reich, H. J. & Murcia, D. A. Stereochemical studies of degenerate silyl rearrangements. Stereospecificity of the tropolone and acetylacetone trialkylsilyl ether rearrangements. *J. Am. Chem. Soc.* **95**, 3418–3420 (1973).
24. Pinnavaia, T. J. & McClarin, J. A. Rearrangements of 1-acetylacetonato-1-methyl-1-silacyclobutane via internal nucleophilic displacement. *J. Am. Chem. Soc.* **96**, 3012–3013 (1974).
25. McClarin, J. A., Schwartz, A. & Pinnavaia, T. J. 1,5-Migrations of silicon between oxygen centers in silyl β -diketones. *J. Organomet. Chem.* **188**, 129–139 (1980).
26. Seeman, J. I. Effect of conformational change on reactivity in organic chemistry. Evaluations, applications, and extensions of Curtin–Hammett Winstein–Holness kinetics. *Chem. Rev.* **83**, 83–134 (1983).
27. Mirarchi, D. & Ritchie, G. L. D. Solution-state conformations of 2-fluoro-, 2-chloro- and 2-bromo-acetophenone: a dipole moment and Kerr effect study. *J. Mol. Struct.* **118**, 303–310 (1984).
28. Kulhánek, J., Böhm, S., Palát, K. Jr & Exner, O. Steric inhibition of resonance: revision of the principle on the electronic spectra of methyl-substituted acetophenones. *J. Phys. Org. Chem.* **17**, 686–693 (2004).
29. Cheng, C. L. & Ritchie, G. L. D. Conformations of the formyl-, acetyl-, and benzoylpyridines. *J. Chem. Soc.* **2**, 1461–1465 (1973).
30. Ramachandran, P. V., Gong, B., Brown, H. C. & Francisco, J. S. Relationship between the structure and enantioselectivity in the asymmetric reduction of 2',6'-disubstituted acetophenones with DIP-chlorideTM. An ab initio study. *Tetrahedr. Lett.* **45**, 2603–2605 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by a grant from the National Institute of General Medical Sciences (GM58133). W.A.C. was supported by a Natural Sciences and Engineering Research Council of Canada Postdoctoral Fellowship. We thank the US National Science Foundation (CRF-0840451) for acquisition of a 400 MHz NMR spectrometer. We thank our colleagues G. Parkin and W. Sattler for an X-ray structure analysis (see the Supplementary Information), and the US National Science Foundation (CHE-0619638) for acquisition of an X-ray diffractometer.

Author Contributions W.A.C. planned and did the vast majority of the experimental work. S.K.R. did the experiments that established the validity of the idea and optimized the allylation of acetylacetone. J.L.L. conceived and directed the project and wrote the manuscript.

Author Information X-ray crystallographic data have been deposited in the Cambridge Crystallographic Data Centre database (<http://www.ccdc.cam.ac.uk/>) under accession code CCDC 874744. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to J.L.L. (leighton@chem.columbia.edu).

Gate-tuning of graphene plasmons revealed by infrared nano-imaging

Z. Fei¹, A. S. Rodin¹, G. O. Andreev¹, W. Bao^{2,3}, A. S. McLeod¹, M. Wagner¹, L. M. Zhang⁴, Z. Zhao², M. Thiemens⁵, G. Dominguez⁶, M. M. Fogler¹, A. H. Castro Neto⁷, C. N. Lau², F. Keilmann⁸ & D. N. Basov¹

Surface plasmons are collective oscillations of electrons in metals or semiconductors that enable confinement and control of electromagnetic energy at subwavelength scales^{1–5}. Rapid progress in plasmonics has largely relied on advances in device nano-fabrication^{5–7}, whereas less attention has been paid to the tunable properties of plasmonic media. One such medium—graphene—is amenable to convenient tuning of its electronic and optical properties by varying the applied voltage^{8–11}. Here, using infrared nano-imaging, we show that common graphene/SiO₂/Si back-gated structures support propagating surface plasmons. The wavelength of graphene plasmons is of the order of 200 nanometres at technologically relevant infrared frequencies, and they can propagate several times this distance. We have succeeded in altering both the amplitude and the wavelength of these plasmons by varying the gate voltage. Using plasmon interferometry, we investigated losses in graphene by exploring real-space profiles of plasmon standing waves formed between the tip of our nano-probe and the edges of the samples. Plasmon dissipation quantified through this analysis is linked to the exotic electrodynamics of graphene¹⁰. Standard plasmonic figures of merit of our tunable graphene devices surpass those of common metal-based structures.

In general, surface plasmons can exist in any material with mobile charge carriers whose response to electric field remains reactive—that is, whose in-plane momentum (q)-dependent and frequency (ω)-dependent complex conductivity, $\sigma(q, \omega) = \sigma_1 + i\sigma_2$, is predominantly imaginary. Of particular interest are plasmons with high momenta, $q_p \gg \omega/c$ (where c is the velocity of light), which may be used for extreme concentration of electromagnetic energy^{1–5}. In conventional

bulk metals, the frequencies of such high- q plasmons reside in the visible or ultraviolet ranges. In graphene, they are expected to appear in the terahertz and infrared domains¹². However, these high- q infrared plasmons are dormant in conventional spectroscopy of graphene. Here we used the scattering-type scanning near-field optical microscope (scattering-type SNOM) to experimentally access high- q plasmons by illuminating the sharp tip of an atomic force microscope (AFM) with a focused infrared beam (Fig. 1a). The momenta imparted by the tip extend up to a few times $1/a$, where $a \approx 25$ nm is the curvature radius of the tip¹³, thus spanning the typical range of infrared plasmon momenta (q_p) in graphene¹². The spatial resolution of scattering-type SNOM is also set by a , and proves to be an order of magnitude smaller than the plasmon wavelength λ_p . The direct observable of our method—the scattering amplitude $s(\omega)$ —is a measure of the electric field strength inside the nanoscale gap between tip and sample. Consequently, the scattering-type SNOM technique enables spectroscopy¹³ and infrared nano-imaging of graphene plasmons without the need to fabricate specialized periodic structures¹⁴. Our imaging data elucidate real-space characteristics of infrared plasmons in graphene such as reflection, interference and damping. All these phenomena can be readily manipulated with gate voltage—a noteworthy property unattainable in metal-based plasmonics.

To probe directly the properties of graphene plasmons, we utilize a frequency $\omega = 892$ cm^{−1} corresponding to a wavelength $\lambda_{\text{IR}} = 11.2$ μm in the infrared (IR) regime where the plasmon is unimpeded by the surface optical phonon supported in graphene/SiO₂/Si structures¹³. The nano-imaging results are shown in Fig. 1b–e, where we plot normalized near-field amplitude $s(\omega) = s_3(\omega)/s_3^{\text{Si}}(\omega)$. Here $s_3(\omega)$ and $s_3^{\text{Si}}(\omega)$ are the

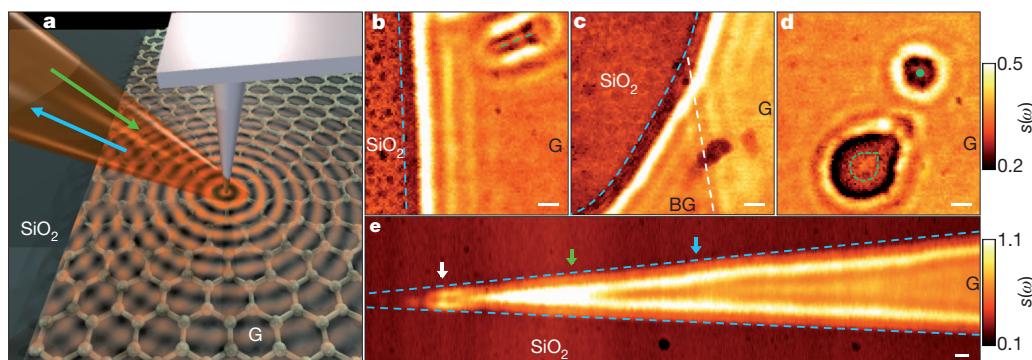


Figure 1 | Infrared nano-imaging experiment and results. **a**, Diagram of an infrared nano-imaging experiment at the surface of graphene (G) on SiO₂. Green and blue arrows display the directions of incident and back-scattered light, respectively. Concentric red circles illustrate plasmon waves launched by the illuminated tip. **b–e**, Images of infrared amplitude $s(\omega = 892$ cm^{−1}) defined in the text taken at zero gate voltage. These images show a characteristic

interference pattern close to graphene edges (blue dashed lines) and defects (green dashed lines and green dot), and at the boundary between single (G) and bilayer (BG) graphene (white dashed line). Additional features marked with arrows in **e** are analysed in refs 27 and 30. Locations of boundaries and defects were determined from AFM topography taken simultaneously with the near-field data. Scale bars, 100 nm. All data were acquired at ambient conditions.

¹Department of Physics, University of California, San Diego, La Jolla, California 92093, USA. ²Department of Physics and Astronomy, University of California, Riverside, California 92521, USA. ³Materials Research Science and Engineering Center, University of Maryland, College Park, Maryland 20742, USA. ⁴Department of Physics, Boston University, Boston, Massachusetts 02215, USA. ⁵Department of Chemistry and Biochemistry, University of California, San Diego, La Jolla, California 92093, USA. ⁶Department of Physics, California State University, San Marcos, California 92096, USA. ⁷Graphene Research Centre and Department of Physics, National University of Singapore, 117542, Singapore. ⁸Max Planck Institute of Quantum Optics and Center for Nanoscience, 85714 Garching, Germany.

third-order demodulated harmonics of the near-field amplitude measured for the given sample and for a Si reference sample, respectively¹³. The near-field amplitude $s(\omega)$ tracks real-space variations in the local electric field underneath the tip, enabling exploration of surface phonon polaritons and surface plasmons^{15,16}.

In Fig. 1b, we present a $s(\omega)$ image acquired at the graphene–SiO₂ interface revealing periodic oscillations of the $s(\omega)$ signal extending along the graphene edge. Point and circular defects (Fig. 1d) trigger circular fringe patterns. Line defects (Fig. 1b) produce elongated, elliptical patterns. Furthermore, we observed fringes at both sides of the boundary between single-layer and bilayer graphene (Fig. 1c). Finally, strongly tapered corners of graphene (Fig. 1e) reveal the two groups of fringes oriented along both edges of graphene in the field of view. In all cases, the periodicity of the fringe patterns is around 100 nm.

Images in Fig. 1 are consistent with the following scenario. Illuminated by focused infrared light, the AFM tip launches plasmon waves of wavelength λ_p propagating radially outward from the tip. Sample edges or defects act as (imperfect) reflectors of the plasmon waves, directing them back to the tip. Therefore complex patterns of interference between launched and reflected plasmons should form inside graphene. We emphasize that our experimental technique does not capture instantaneous snapshots of these complex patterns. Instead, whereas the tip ‘launches’ plasmon waves propagating in all directions, it only ‘detects’ the cumulative near-field plasmonic signal arising underneath it. This stands in rough analogy with the operating principle of sonar echolocation. In the top panels of Fig. 2a, we sketch the plasmon interference pattern in the form of plasmon amplitude, revealing standing wave oscillations between the tip and sample edge. As the tip is scanned towards the edge, it registers these oscillations with periodicity given by $\lambda_p/2$, as shown in the bottom panels of Fig. 2a. Our plasmon interference interpretation is supported by theoretical estimates of the wavelength λ_p . The plasmon dispersion of a two-dimensional metal residing at the interface of vacuum (dielectric constant $\epsilon_0 = 1$) and a substrate with dielectric function $\epsilon_{\text{sub}}(\omega)$ is given by the formula:

$$q_p = \frac{i\omega\kappa}{2\pi\sigma} \quad (1)$$

where $\kappa(\omega) = \kappa_1 + i\kappa_2 \equiv [\epsilon_0 + \epsilon_{\text{sub}}(\omega)]/2$ represents the complex effective dielectric function of the environment for graphene. Assuming that the conductivity (σ) of graphene takes a Drude form with relaxation time τ , equation (1) can be rewritten as:

$$q_p = \frac{\hbar^2 \kappa(\omega)}{2e^2 E_F} \omega \left(\omega + \frac{i}{\tau} \right) \quad (2)$$

where \hbar is the reduced Planck’s constant, e is the elementary charge, and E_F is the Fermi energy. Derivation of these equations and further refinements are discussed in the Supplementary Information. The real part of $q_p = q_1 + iq_2$ determines the plasmon wavelength $\lambda_p = 2\pi/q_1$, and the ratio between the imaginary part and the real part defines the plasmon damping rate $\gamma_p = q_2/q_1$. In graphene, the Dirac-like dispersion of the Fermi energy $E_F = \hbar v_F k_F$ with Fermi momentum $k_F = \sqrt{\pi|n|}$ (ref. 8) implies $|n|^{-1/2}$ scaling of the plasmon momentum with the carrier density n at fixed ω . Here $v_F \approx c/300 = 10^6 \text{ m s}^{-1}$ is the Fermi velocity. Finally, using frequency $\omega = 892 \text{ cm}^{-1}$ and $n \approx 8 \times 10^{12} \text{ cm}^{-2}$ determined from the micro-Raman probe (see below) at the graphene edge, we find $\lambda_p \approx 200 \text{ nm}$ from equation (2), which is roughly twice the distance between fringes in Fig. 1b–e.

The images in Fig. 1b–e contain rich insights into processes governing plasmon propagation and losses on the surface of graphene. It is therefore instructive to examine line profiles along the direction normal to the sample edges. In Fig. 2b we show a plot obtained by averaging 150 such profiles—a procedure used to improve the signal-to-noise ratio. We find that the fringe widths increase from the interior to the edge of graphene, implying that the plasmon wavelength λ_p likewise increases. This behaviour is due to enhancement of the carrier

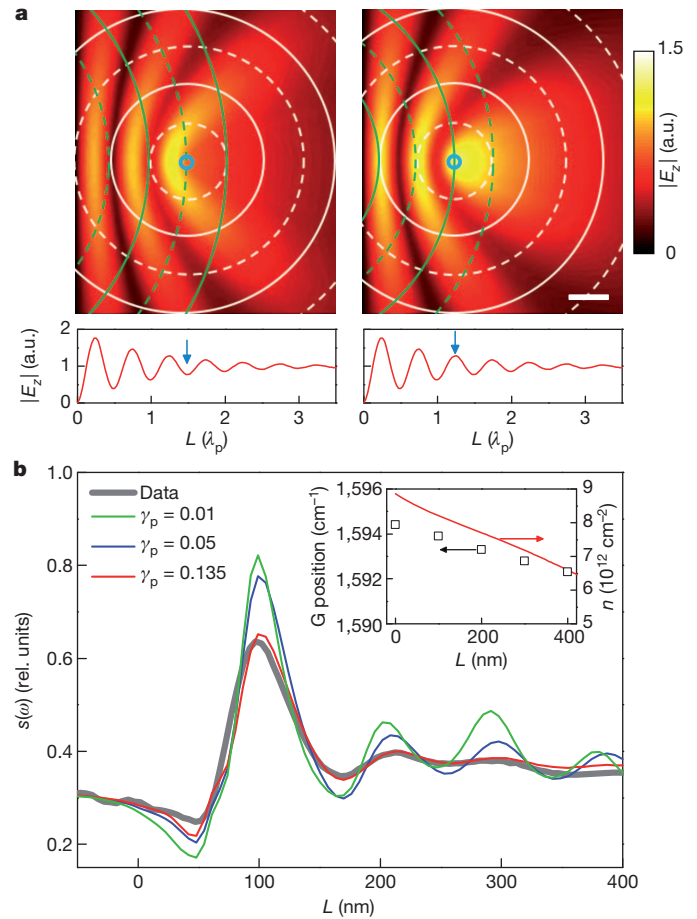


Figure 2 | Spatial variation of the electric field and near-field amplitude at the graphene edge. For all panels, graphene is present at $L > 0$, and SiO₂ not covered by graphene is displayed at $L < 0$. **a**, Upper panels, illustration of interference between tip-launched plasmon waves (white) and their reflection (green) from the edge at $L = 0$. Solid and dashed lines correspond to positive and negative field maxima of the propagating plasmon, respectively. False colour plots of the absolute value of electric field $|E_z|$ reveal standing waves formed between the tip and the edge. Left and right panels show snapshots of destructive (minimum signal) and constructive (maximum signal) interference underneath the tip, respectively. Scale bar, $0.5\lambda_p$. Lower panels, profiles of $|E_z|$ underneath the tip versus its distance to the edge. The blue circles and arrows mark the positions of the tip. **b**, Experimental (grey) and calculated (colour) $s(\omega)$ line profiles at zero gate voltage. Inset, G-peak positions inferred from micro-Raman data (squares) and the carrier density profile (red line) we used to model the plasmonic standing wave (in **b**). The Raman G-peak positions are associated with the variation of the local carrier density in graphene (n : right-hand scale)^{17,18,29}.

density n near the sample edge, which is verified by our micro-Raman experiments (Fig. 2b inset)^{17,18}. Thus, plasmonic interference patterns reported in Fig. 2b uncover the usefulness of infrared imaging for the nanoscale determination of local carrier density in graphene. In Fig. 2b we also show modelling results of plasmon profiles following a procedure detailed in the Supplementary Information. Our modelling provides a quantitative account of plasmon interferometry data. The carrier density profile (red curve in Fig. 2b inset) and the damping rate γ_p constitute the adjustable parameters of the model. Because plasmons in our experiments are launched and detected by the same point source (the AFM tip), the interference amplitude necessarily exhibits decay from the sample edge even when the damping rate is assumed to be vanishingly small (green trace in Fig. 2b). The best fit to the amplitude profile is achieved for significantly stronger damping, with $\gamma_p = 0.135$.

According to equations (1) and (2), the plasmon wavelength λ_p is directly determined by the carrier density n . We demonstrate this

unique aspect of graphene plasmonics experimentally through imaging under gate bias (Fig. 3a). Over a range of V_g (gate voltage) values from +30 V to −20 V, the hole density n in our samples increases monotonically, a consequence of significant unintentional hole doping present even in ungated graphene/SiO₂/Si structures (Fig. 2b inset). This tuning of carrier density produces systematic variations in the plasmonic profiles: fringe amplitude and periodicity are both enhanced with increasing n . By inferring λ_p directly from the fringe width, we observe a systematic decrease in λ_p with the reduction in hole density. Our gate-dependent data for λ_p approximately follow the $\lambda_p \propto |n|^{1/2}$ law predicted for monolayer graphene¹⁹. In contrast, the plasmon damping rate does not show clear gate dependence and is roughly equal to 0.135 ± 0.015 at all V_g . This magnitude of γ_p significantly exceeds theoretical estimates for graphene with similar electronic mobility, $\mu \approx 8,000 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$ (ref. 12).

It is important to understand why plasmon damping in our structures is abnormally strong. According to equations (1) and (2), two additive contributions define damping rate as $\gamma_p \approx (\sigma_1/\sigma_2) + (\kappa_2/\kappa_1)$. The first term is associated with plasmonic losses inherent in graphene, whereas the second term describes losses due to the SiO₂ substrate. At $\omega = 892 \text{ cm}^{-1}$, we estimate $\kappa = 2.52 + 0.13i$ and hence, $\kappa_2/\kappa_1 \approx 0.05$, based on our ellipsometric measurements of SiO₂/Si wafers. The

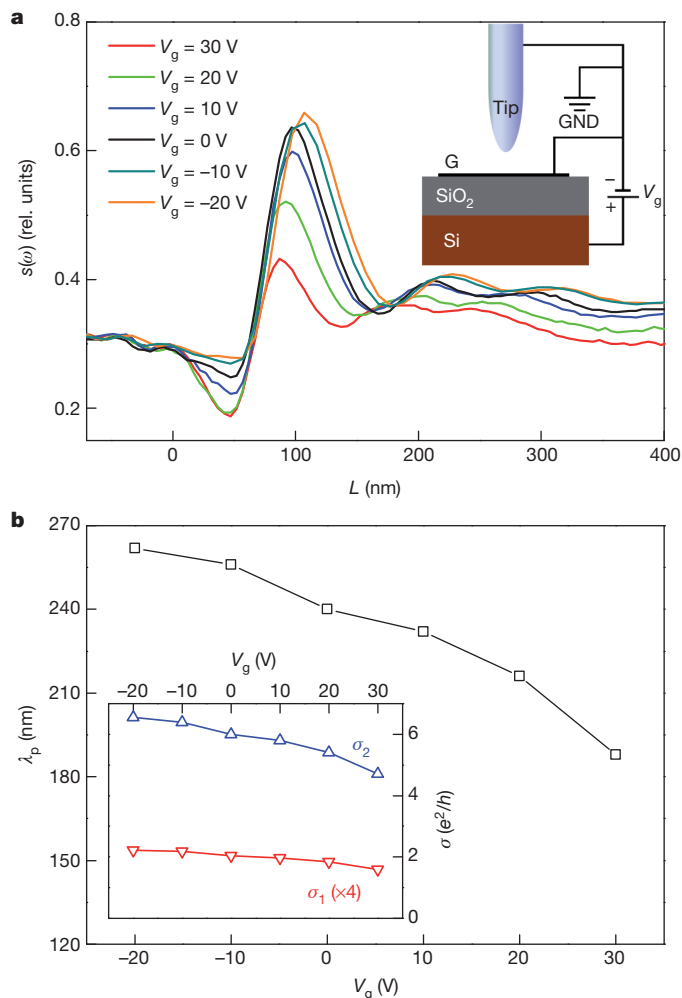


Figure 3 | Electrostatically tunable plasmons in back-gated graphene. **a**, Line profiles of $s(\omega)$ perpendicular to the graphene edge at various gate voltages, V_g . Inset, our gate bias set-up. Both the metalized tip and graphene are at ground (GND) potential. **b**, Gate-dependent plasmon wavelength λ_p . Inset, σ_1 and σ_2 of graphene at various gate voltages, estimated from λ_p and γ_p , as described in the text. These conductivity data show directly that the infrared response of graphene is predominantly reactive, $\sigma_2 \gg \sigma_1$, thus fulfilling an essential precondition for excitation of plasmons.

resulting value of $\sigma_1/\sigma_2 \approx 0.08$ is unexpectedly high, being three to four times higher than the estimate of $\sigma_1/\sigma_2 = (\omega\tau)^{-1}$ obtained from equation (2) using the relaxation rate $\tau^{-1} \approx 20 \text{ cm}^{-1}$, corresponding to a typical d.c. mobility of our samples. This discrepancy has two possible interpretations. Excessive losses may originate from an enhanced electronic relaxation rate at infrared frequencies compared to that established in d.c. transport. Alternatively, losses may be unrelated to free carrier mobility/dynamics and may instead be associated with extrinsic factors such as surface irregularities. Our plasmonic interferometry data in Figs 2 and 3 provide strong support for the former hypothesis. Indeed, these images yield γ_p and λ_p (Fig. 3b) and thus allow us to determine the complex optical conductivity of graphene (Fig. 3b inset) based on the following formulae: $\sigma_2 \approx \frac{\omega\kappa}{4\pi^2} \lambda_p$; $\sigma_1 = \sigma_2(\gamma_p - \kappa_2/\kappa_1)$. We remark that these relations between plasmonic parameters (λ_p , γ_p) and the complex optical conductivity σ of graphene hold true for any plasmonic material for which $\sigma_1 \ll \sigma_2$. Therefore these expressions apply even if the frequency-dependent conductivity deviates from the simple Drude model. The appeal of this analysis lies in establishing a link between real-space plasmonic profiles and the optical constants inferred from conventional infrared spectroscopy.

Our imaging data reveal that the real part of the conductivity of graphene is as high as $\sigma_1 \approx 0.5e^2/h$. This value greatly exceeds the theoretical estimate of σ_1 ($\omega = 892 \text{ cm}^{-1}$, $n = 8 \times 10^{12} \text{ cm}^{-2}$, $\tau^{-1} = 20 \text{ cm}^{-1}$) = $0.13e^2/h$ obtained from a model of non-interacting Dirac quasiparticles weakly scattered by disorder^{20–22}, but is comparable to infrared spectroscopy results for back-gated graphene on SiO₂/Si (ref. 10). In an ideal, non-interacting graphene, $\sigma_1(\omega < 2E_F)$ is vanishingly small owing to the phenomenon of ‘Pauli blocking’¹⁰. Thus the source of strong plasmonic losses in our back-gated samples is traced to the unexpectedly large magnitude of σ_1 . This result supports the notion of prominent many-body effects in graphene beyond the picture of non-interacting Dirac fermions^{23–25}. Further experiments on suspended graphene, as well as devices using various types of dielectric substrates (such as hexagonal boron nitride), are needed to disentangle the roles of electron–electron and graphene–substrate interactions in the dissipation we observe at infrared frequencies. Our work uncovers an experimental path and analysis methodology for these future studies of many-body interactions in graphene.

Infrared nano-imaging experiments reported here have established graphene/SiO₂/Si structures as a potent plasmonic medium that enables voltage control of both the wavelength and the amplitude of the plasmons. Higher gate voltages than used in our study will allow for the binary on/off switching of plasmon propagation with a possibility of potentially local control by a top gate or a biased tip. The plasmon wavelength in graphene, $\lambda_p \approx 200 \text{ nm}$, is one of the shortest imaged for any material, whereas the propagation length is on a par with Au in experiments monitoring strongly confined plasmons launched by AFM tips¹⁶. An important figure of merit, $\lambda_{\text{IR}}/\lambda_p = 50\text{--}60$, for our back-gated devices surpasses that of conventional Ag-based structures¹². Intrinsic plasmonic losses in graphene that we analysed in detail can be substantially reduced or even eliminated through population inversion²⁶. We stress that plasmon tuning is realized here in the architecture of a metal-oxide-semiconductor device (we used graphene-oxide-silicon), which is a ubiquitous system in modern information processing. Furthermore, the performance of even the first generation of plasmonic devices reported here and in ref. 27 is rather promising compared to non-tunable metal-based structures¹². For all these reasons, we believe that graphene may be an ideal medium for active infrared plasmonics.

METHODS SUMMARY

Infrared nano-imaging. Our scattering-type SNOM apparatus (Neaspec) is based on an AFM operating in tapping mode. Measurements were performed at an AFM tapping frequency of $\Omega = 270 \text{ kHz}$ and a tapping amplitude of 40 nm. The shaft of the conducting AFM tip acts as an antenna that boosts the efficiency of

near-field interaction²⁸. The back-scattered signal registered by the detector is strongly dependent on the tip-sample distance. This enables isolation of the genuine near-field contribution from the overall back-scattered signal, which is periodically modulated at harmonics of the tapping frequency Ω . The m th harmonic component of this signal $s_m \exp(i\phi_m)$, termed the m th demodulated signal (here $m = 3$), represents the desired near-field contribution.

Samples and devices. Our graphene samples were obtained by mechanical cleavage of bulk graphite and then transferred to SiO₂/Si substrate. To avoid surface contamination by lithographic procedures, bulk graphite connected to our graphene sample was used as an electrode in our back-gating experiments. To verify the gating functionality of our devices, we first performed a spectroscopic study of the hybrid plasmon-phonon resonance at various gate voltages and found good agreement with the published data¹³. Transport measurements of graphene samples fabricated following identical procedures indicate that the typical mobility of our graphene samples is about 8,000 cm² V⁻¹ s⁻¹. Plasmon imaging experiments were completed for more than 30 graphene samples. All these structures exhibited highly reproducible behaviour and consistent values for the plasmon wavelength and damping. The data displayed in Figs 1–3 were obtained for devices with some of the weakest damping, revealing the largest number of plasmonic oscillations. Nevertheless, even in these devices the plasmonic losses are stronger than expectations based on typical electronic mobility measurements.

Micro-Raman measurements. According to previous studies, the position of the G-peak in the Raman spectrum of graphene is directly linked to its carrier density^{17,18,29}. Therefore, the G-peak profile shown in Fig. 2b inset reflects the range of the variation in graphene carrier density close to the edge^{17–19}. Our micro-Raman experiments were carried out using a Renishaw inVia Raman microscope equipped with a 50 \times , NA = 0.75 long-distance objective, a 1,800 lines per mm grating and an XY stage with a resolution of 100 nm. The spot size in these experiments is limited by diffraction. Therefore, the fragment of the line profile of the G-peak frequency shown in Fig. 2b inset is instrumentally broadened.

Received 29 February; accepted 17 May 2012.

Published online 20 June 2012.

- Atwater, H. A. The promise of plasmonics. *Sci. Am.* **296**, 56–62 (2007).
- West, P. R. *et al.* Searching for better plasmonic materials. *Laser Photon. Rev.* **4**, 795–808 (2010).
- Stockman, M. I. Nanoplasmonics: the physics behind the applications. *Phys. Today* **64**, 39–44 (2011).
- Maier, S. A. *Plasmonics: Fundamentals and Applications* Ch. 4 (Springer, 2007).
- Schuller, J. A. *et al.* Plasmonics for extreme light concentration and manipulation. *Nature Mater.* **9**, 193–204 (2010).
- Nagpal, P., Lindquist, N. C., Oh, S.-H. & Norris, D. J. Ultrasoft patterned metals for plasmonics and metamaterials. *Science* **325**, 594–597 (2009).
- Lal, S., Link, S. & Halas, N. J. Nano-optics from sensing to waveguiding. *Nature Photon.* **1**, 641–648 (2007).
- Castro Neto, A. H., Guinea, F., Peres, N. M. R., Novoselov, K. S. & Geim, A. K. The electronic properties of graphene. *Rev. Mod. Phys.* **81**, 109–162 (2009).
- Wang, F. *et al.* Gate-variable optical transitions in graphene. *Science* **320**, 206–209 (2008).
- Li, Z. Q. *et al.* Dirac charge dynamics in graphene by infrared spectroscopy. *Nature Phys.* **4**, 532–535 (2008).
- Vakil, A. & Engheta, N. Transformation optics using graphene. *Science* **332**, 1291–1294 (2011).
- Jablan, M., Buljan, H. & Soljacic, M. Plasmonics in graphene at infrared frequencies. *Phys. Rev. B* **80**, 245435 (2009).
- Fei, Z. *et al.* Infrared nanoscopy of Dirac plasmons at the graphene-SiO₂ interface. *Nano Lett.* **11**, 4701–4705 (2011).
- Ju, L. *et al.* Graphene plasmonics for tunable terahertz metamaterials. *Nature Nanotechnol.* **6**, 630–634 (2011).
- Huber, A., Ocelic, N., Kazantsev, D. & Hillenbrand, R. Near-field imaging of mid-infrared surface phonon polariton propagation. *Appl. Phys. Lett.* **87**, 081103 (2005).
- Dallapiccola, R., Dubois, C., Gopinath, A., Stellacci, F. & Dal Negro, L. Near-field excitation and near-field detection of propagating surface plasmon polaritons on Au waveguide structures. *Appl. Phys. Lett.* **94**, 243118 (2009).
- Casiraghi, C. *et al.* Raman spectroscopy of graphene edges. *Nano Lett.* **9**, 1433–1441 (2009).
- Caridad, J. M. *et al.* Effects of particle contamination and substrate interaction on the Raman response of unintentionally doped graphene. *J. Appl. Phys.* **108**, 084321 (2010).
- Hwang, E. H. & Das Sarma, S. Dielectric function, screening, and plasmons in two-dimensional graphene. *Phys. Rev. B* **75**, 205418 (2007).
- Ando, T., Zheng, Y. & Suzuura, H. Dynamical conductivity and zero-mode anomaly in honeycomb lattices. *J. Phys. Soc. Jpn* **71**, 1318–1324 (2002).
- Peres, N. M. R., Guinea, F. & Castro Neto, A. H. Electronic properties of disordered two-dimensional carbon. *Phys. Rev. B* **73**, 125411 (2006).
- Gusynin, V. P. & Sharapov, S. G. Transport of Dirac quasiparticles in graphene: Hall and optical conductivities. *Phys. Rev. B* **73**, 245411 (2006).
- Grushin, A. G., Valenzuela, B. & Vozmediano, M. A. H. Effect of Coulomb interactions on the optical properties of doped graphene. *Phys. Rev. B* **80**, 155417 (2009).
- Peres, N. M. R., Ribeiro, R. M. & Castro-Neto, A. H. Excitonic effects in the optical conductivity of gated graphene. *Phys. Rev. Lett.* **105**, 055501 (2010).
- Hwang, J., Leblanc, J. P. F. & Carbotte, J. P. Optical self-energy in graphene due to correlations. *J. Phys. Condens. Matter* **24**, 245601 (2012).
- Rana, F. Graphene terahertz plasmon oscillators. *IEEE Trans. Nanotechnol.* **7**, 91–99 (2008).
- Chen, J. *et al.* Optical nano-imaging of gate-tunable graphene plasmons. *Nature* <http://dx.doi.org/10.1038/nature11254> (this issue).
- Novotny, L. & Hecht, B. *Principles of Nano-Optics* (Cambridge Univ. Press, 2006).
- Das, A. *et al.* Monitoring dopants by Raman scattering in an electrochemically top-gated graphene transistor. *Nature Nanotechnol.* **3**, 210–215 (2008).
- Rodin, A. S. *Electronic Properties of Low-Dimensional Systems*. PhD thesis, UCSD (2012).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We acknowledge support from AFOSR, ONR and DARPA. The analysis of plasmonic losses and many-body effects was supported by DOE-BES grant DE-FG02-00ER45799. W.B., Z.Z. and C.N.L. were supported by NSF DMR/1106358, ONR N00014-09-1-0724, ONR/DMEA H94003-10-2-1003 and FENA Focus Center. G.D. and M.T. were supported by NASA. M.M.F. was supported by UCOP and NSF PHY11-25915. A.H.C.N. acknowledges NRF-CRP grant R-144-000-295-281. L.M.Z. was supported by DOE grant no. DE-FG02-08ER46512. M.W. thanks the Alexander von Humboldt Foundation for financial support. F.K. was supported by Deutsche Forschungsgemeinschaft through the Cluster of Excellence Munich Centre for Advanced Photonics.

Author Contributions All authors were involved in designing the research, performing the research, and writing the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/nature. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to D.N.B. (dbasov@physics.ucsd.edu).

The diet of *Australopithecus sediba*

Amanda G. Henry¹, Peter S. Ungar^{2,3}, Benjamin H. Passey⁴, Matt Sponheimer^{3,5}, Lloyd Rossouw^{6,7,8}, Marion Bamford⁹, Paul Sandberg⁵, Darryl J. de Ruiter^{3,10} & Lee Berger³

Specimens of *Australopithecus sediba* from the site of Malapa, South Africa (dating from approximately 2 million years (Myr) ago)¹ present a mix of primitive and derived traits that align the taxon with other *Australopithecus* species and with early *Homo*². Although much of the available cranial and postcranial material of *Au. sediba* has been described^{3–6}, its feeding ecology has not been investigated. Here we present results from the first extraction of plant phytoliths from dental calculus of an early hominin. We also consider stable carbon isotope and dental microwear texture data for *Au. sediba* in light of new palaeoenvironmental evidence. The two individuals examined consumed an almost exclusive C₃ diet that probably included harder foods, and both dicotyledons (for example, tree leaves, fruits, wood and bark) and monocotyledons (for example, grasses and sedges). Like *Ardipithecus ramidus* (approximately 4.4 Myr ago) and modern savanna chimpanzees, *Au. sediba* consumed C₃ foods in preference to widely available C₄ resources. The inferred consumption of C₃ monocotyledons, and wood or bark, increases the known variety of early hominin foods. The overall dietary pattern of these two individuals contrasts with available data for other hominins in the region and elsewhere.

Early hominin diet is central to the study of human origins. Dietary data come from a variety of sources that provide different information about the foods consumed. Carbon isotopes indicate whether an animal ate C₃ resources (for example, trees, shrubs, some herbs, and animals eating these plants), C₄ resources (for example, most tropical grasses, sedges, and animals eating these plants) or a combination of these⁷. Chimpanzees consume C₃ plants such as fruits and leaves even when C₄ grasses are abundant^{8,9}. By contrast, isotopic evidence indicates that *Australopithecus*, *Paranthropus* and early *Homo* consumed variable amounts of C₄ foods, but their diets included more C₄ foods than the diets of modern chimpanzees, indicating that they probably used their environments in different ways than do living apes^{10–12}. Dental microwear provides information about the fracture properties of foods consumed shortly before the death of an animal¹³. Previous studies of *Australopithecus africanus*, and particularly of *Australopithecus afarensis* and *Paranthropus boisei*, suggest limited variance in food hardness, whereas *Homo erectus* and *Paranthropus robustus* have greater variance in texture complexity consistent with a more diverse or variable diet^{14–17}. Phytoliths are plant-produced silica bodies that take on taxon-specific morphology¹⁸. They can become trapped in dental calculus and can therefore provide a record of the plant foods an individual consumed. Phytoliths have been recovered from Neanderthal and modern human calculus¹⁹, but not yet from earlier hominins.

Palaeoenvironmental evidence from Malapa provides a context for interpreting the dietary proxy data. Pollen, phytoliths and wood fragments from *Podocarpus* trees and other woody taxa recovered from a probable carnivore coprolite indicate a forest or woodland

biome in the vicinity²⁰. Sediment samples from the hominin layer (Facies D²¹) preserve many grass phytoliths, predominantly long-necked bilobates and saddles typical of C₄ grasses in summer rainfall areas, consistent with a more open savanna or grassland biome (Supplementary Text, Supplementary Table 4 and Supplementary Fig. 6). The presence of grazers, including *Equus* and *Megalotragus*, supports the interpretation of abundant grass around Malapa²¹. The Malapa bovid fossils have carbon isotope compositions indicating nearly pure C₄ diets, and the rodent fossils have carbon isotope compositions demonstrating high C₄ consumption, as is common among extant rodents around the site (Table 1 and Supplementary Table 1). Thus, the palaeoenvironment probably included areas with abundant grass and woody vegetation.

The carbon isotope compositions of *Au. sediba* individuals MH1 and MH2 were measured using laser ablation isotope ratio mass spectrometry²². The $\delta^{13}\text{C}$ values of both individuals fall outside the range of all 81 previously analysed African early hominins, indicating a nearly pure C₃ diet (Fig. 1, Table 1 and Supplementary Tables 1 and 2)^{10–12,23–26}. The $\delta^{13}\text{C}$ values of MH1 and MH2 are more than two standard deviations below the mean of *Au. africanus*^{11,24}, and are closest to (but still below) the range of much older (4.4 Myr ago) *Ardipithecus ramidus*²⁶. Their values are comparable to those of savanna chimpanzees (corrected for anthropogenic changes in the $\delta^{13}\text{C}$ of atmospheric CO₂)^{8,9}, but higher than those of chimpanzees living in forest environments²⁷. Thus, the carbon isotope compositions of MH1 and MH2 are unusual for hominins, and more typical of C₃ specialists like giraffes (Supplementary Table 1)^{25,26}. If these individuals are representative of the species, *Au. sediba* had a diet that was different from those of most early African hominins studied so far.

Molars of both *Au. sediba* individuals were analysed using dental microwear texture analysis. Two measures of surface texture, area-scale fractal complexity and exact proportion length-scale anisotropy of relief, are useful for distinguishing a broad variety of mammals on the basis of food hardness and toughness, respectively¹⁴. The values for anisotropy are well within the ranges of most early hominins (Fig. 2 and Supplementary Table 3), but the complexity values are higher than those of most specimens of *Australopithecus* spp., *Homo habilis* and *P. boisei*. The complexity value for MH2 falls within the upper end of the *Au. africanus* range, but the juvenile male MH1 has a much higher value, within the range of only *P. robustus*, suggesting the consumption of hard objects before death. The complexity values for both *Au. sediba* specimens are between the medians for *H. erectus* and *P. robustus*, and the difference between them approximates the interquartile ranges of these two hominin taxa.

Thirty-eight phytoliths were recovered from dental calculus on two teeth of MH1, including 15 dicotyledon morphotypes, 9 monocotyledon morphotypes and 14 indeterminate forms (Table 2, Fig. 3 and Supplementary Table 5). This assemblage differs considerably from

¹Plant Foods and Hominin Dietary Ecology Research Group, Max Planck Institute for Evolutionary Anthropology, Deutscher Platz 6, 04103 Leipzig, Germany. ²Department of Anthropology, University of Arkansas, Old Main 330, Fayetteville, Arkansas 72701, USA. ³The Institute for Human Evolution, School of Geosciences, University of the Witwatersrand, Private Bag 3, WITS 2050, Johannesburg, South Africa. ⁴Department of Earth and Planetary Sciences, Johns Hopkins University, 3400 North Charles Street, Baltimore, Maryland 21218, USA. ⁵Department of Anthropology, University of Colorado at Boulder, 233 UCB, Boulder, Colorado 80309-0233, USA. ⁶Department of Archaeology, National Museum Bloemfontein, PO Box 266, Bloemfontein 9300, South Africa. ⁷The Department of Plant Sciences, University of the Free State, PO Box 339, Bloemfontein 9300, South Africa. ⁸The Department of Biology, University of Washington, Box 351800, 24 Kincaid Hall, Seattle, Washington 98195-1800, USA. ⁹BPI Palaeontology, School of Geosciences, University of the Witwatersrand, Private Bag 3, WITS 2050, Johannesburg, South Africa. ¹⁰Department of Anthropology, Texas A&M University, College Station, Texas 77843-4352, USA.

Table 1 | Stable carbon and oxygen isotope compositions of Malapa fauna, MH1 and MH2

Specimen	Taxon	$\delta^{13}\text{C}$ (‰, VPDB)	$\delta^{13}\text{C}$ adj* (‰, VPDB)	$\delta^{18}\text{O}$ (‰, SMOW)
MB1 analysis 1	Bovidae	0.3	0.8	26.0
MB1 analysis 2	Bovidae	0.4	0.9	26.4
MB1 mean	Bovidae	0.4	0.9	26.2
MB2 analysis 1	Bovidae	-0.7	-0.2	25.9
MB2 analysis 2	Bovidae	-0.9	-0.4	25.2
MB2 mean	Bovidae	-0.8	-0.3	25.6
Bovidae mean	Bovidae	-0.2	0.3	25.9
Bovidae s.d.	Bovidae	0.8	0.8	0.4
MR1	Rodentia	-4.5	-4.0	21.4
MR2 analysis 1	Rodentia	-5.4	-4.9	19.8
MR2 analysis 2	Rodentia	-5.7	-5.2	20.1
MR2 mean	Rodentia	-5.6	-5.1	20.0
MR3 analysis 1	Rodentia	-5.0	-4.5	18.8
MR3 analysis 2	Rodentia	-5.3	-4.8	18.9
MR3 mean	Rodentia	-5.2	-4.7	18.9
Rodentia mean	Rodentia	-5.1	-4.6	20.1
Rodentia s.d.	Rodentia	0.6	0.6	1.3
Modern <i>Otomys</i> mean**	Rodentia		-5.5	21.5
Modern <i>Otomys</i> s.d.**	Rodentia		2.0	4.9
MH1 analysis 1	<i>Au. sediba</i>	-12.8	-12.3	20.1
MH1 analysis 2	<i>Au. sediba</i>	-11.5	-11.0	20.6
MH1 mean	<i>Au. sediba</i>	-12.2	-11.7	20.4
MH2 analysis 1	<i>Au. sediba</i>	-12.5	-12.0	18.2
MH2 analysis 2	<i>Au. sediba</i>	-11.7	-11.2	18.4
MH2 mean	<i>Au. sediba</i>	-12.1	-11.6	18.3
<i>Australopithecus sediba</i> mean	<i>Au. sediba</i>	-12.2	-11.7	19.4
<i>Australopithecus sediba</i> s.d.	<i>Au. sediba</i>	0.1	0.1	1.5

SMOW, Standard Mean Ocean Water; VPDB, Vienna Pee Dee belemnite.

* $\delta^{13}\text{C}$ values obtained through laser ablation are lower than those obtained through conventional acid hydrolysis by about 0.5‰ (ref. 22). Values have been adjusted in this column accordingly.

** Modern values have been adjusted to compensate for the fossil-fuel effect. Five *Otomys* individuals (which are typically grass consumers and potential analogues for the fossil rodents analysed from Malapa) have been sampled from an active owl roost near the site. Most other modern rodents from the roost consumed far more C_3 vegetation, and as a result the overall roost mean ($n = 24$; $\delta^{13}\text{C}$, $-9.8 \pm 5.3\text{‰}$; $\delta^{18}\text{O}$, $19.8 \pm 4.3\text{‰}$) is much lower than that of the available fossil sample (Supplementary Table 1).

those recovered from the Facies D sediment, which contains abundant C_4 grass phytoliths, and from the breccia block containing MH1, in which no phytoliths were found (Supplementary Text, Supplementary Fig. 6 and Supplementary Table 4). Unlike the sediment, there were no distinctive C_4 monocotyledon morphotypes in the calculus. These monocotyledon phytoliths most probably represent the consumption of tropical shade- and water-loving C_3 grasses and sedges. The presence of fruit, leaf and wood or bark phytoliths suggests that all of these were consumed to some extent, although quantification of their relative importance is hampered by differences in phytolith density among taxa and plant parts¹⁹. Bark and woody tissues have

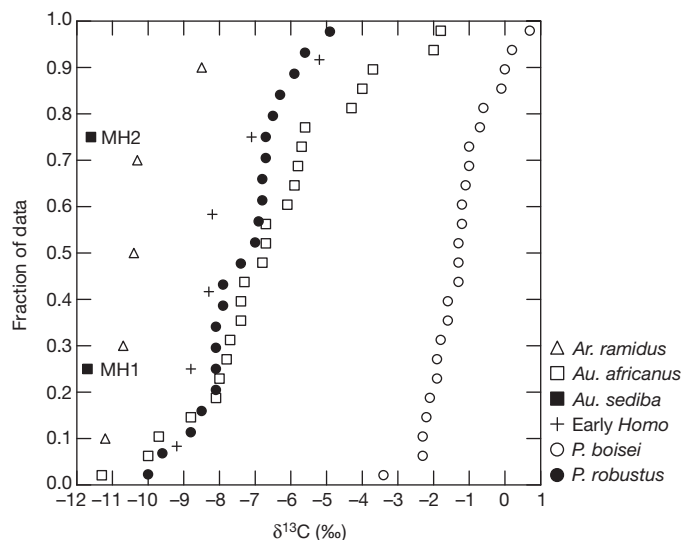


Figure 1 | $\delta^{13}\text{C}$ values for early hominins. *Au. sediba* individuals MH1 and MH2 have $\delta^{13}\text{C}$ values ($\delta^{13}\text{C} = ({}^{13}\text{C}/{}^{12}\text{C}_{\text{sample}} / {}^{13}\text{C}/{}^{12}\text{C}_{\text{standard}} - 1) \times 1000$) outside the range of previously sampled hominins, and indicate that these individuals consumed almost exclusively C_3 foods. Values for other hominin specimens have been collected from the literature and are presented alongside the data for the Malapa modern and fossil taxa in Supplementary Table 1.

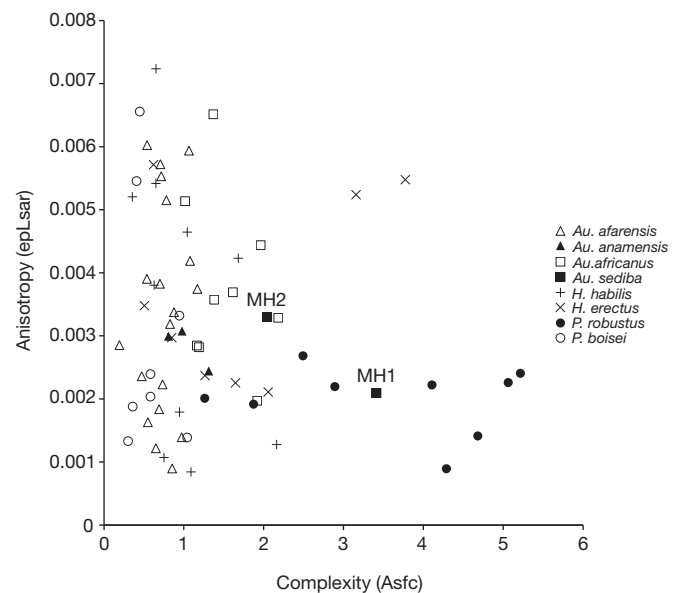


Figure 2 | Dental microwear patterns for individuals MH1 and MH2 compared to those of other fossil hominins. Area-scale fractal complexity and exact proportion length-scale anisotropy of relief typically differentiate diets based on food hardness and toughness, respectively. Although the anisotropy values of *Au. sediba* individuals MH1 and MH2 fall within the range of most early African hominins, the complexity values suggest MH1 and MH2 consumed hard foods, similar to *H. erectus* and *P. robustus*.

Table 2 | Counts of phytoliths recovered from the calculus of MH1

Tooth	Dicotyledon			Monocotyledon			Indeterminate	Totals
	Leaf	Fruit	Wood/Bark	Sedge	Grass	Palm		
LI ²	1	1	7	2	1	1	11	24
LP ³	2	–	4	2	3	–	3	14
Totals	3	1	11	4	4	1	14	38

LI², upper left second incisor; LP³, upper left first premolar

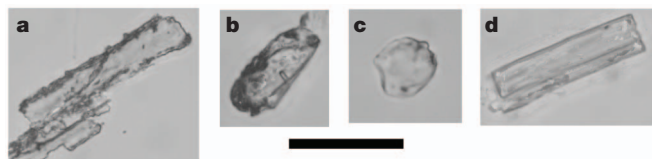


Figure 3 | Phytoliths from the dental calculus of MH1. **a**, A dicotyledon fruit phytolith. **b**, A dicotyledon wood or bark phytolith. **c**, A small bulliform phytolith from grass. **d**, A sedge phytolith. Scale bar, 50 μ m.

not been documented previously as a dietary component for hominins, although these foods are consumed by many primates and contain proteins and soluble sugars^{28,29}. The diversity in phytolith types, despite the small available sample, suggests a varied diet.

The three lines of evidence suggest these two individuals had an unexpected diet compared to those of African hominins of a similar antiquity. Dental microwear texture analysis suggests consumption of at least some hard foods to an extent not previously seen in most *Australopithecus* spp. or in *P. boisei*, apparently more like that of *H. erectus* and *P. robustus*. Carbon isotope data indicate that these individuals were nearly exclusive C_3 feeders, much like today's savanna chimpanzees. The recovered phytoliths indicate that C_3 dicotyledons and monocotyledons were probably a considerable component of the diet, and the variety of morphotypes implies that MH1 consumed a wide variety of foods. Among the early African hominins, the closest match is arguably *Ar. ramidus* based on their similar carbon isotope compositions, although microwear texture data and phytoliths are unavailable for this taxon.

All three lines of evidence are consistent with *Au. sediba* having made use of habitats in which C_3 dicotyledons and monocotyledons grew within a regional environment with abundant C_4 grasses, similar to a gallery forest. Like *Ar. ramidus* and savanna chimpanzees, MH1 and MH2 seem to have consumed C_3 foods in preference to widely available C_4 resources, and they probably included bark and other fracture-resistant foods as at least a seasonal part of their diet. This interpretation is consistent with postcranial evidence suggesting that *Au. sediba* was a facultative tree climber⁶, possibly indicating a reliance on arboreal resources. However, our results raise several questions about other aspects of *Au. sediba* behaviour. For example, the carbon isotope data suggest that these individuals were limited to C_3 resources, which could indicate that these individuals required large home ranges to acquire preferred fruit resources, like savanna chimpanzees. However, if they consumed a wider variety of C_3 resources during periods of fruit scarcity, as evinced by the phytoliths, this might have reduced the need for larger home ranges.

Previous analyses have shown that *Au. sediba* has an unusual suite of morphological features^{2–6}, and our results present new oddities and questions. This work and other recent studies suggest that the diets of australopiths were diverse³⁰ (both within and between most species), which mirrors recently recognized morphological diversity among the australopiths². Nevertheless, our interpretations of the diet of *Au. sediba* must be tempered by the small sample that is available. We expect that studying additional individuals from the site will provide a better understanding of the dietary ecology of *Au. sediba*.

METHODS SUMMARY

Carbon and oxygen isotope ratios of the specimens (upper right first incisor (RI¹) of MH1 and upper right second molar (RM²) of MH2) were determined using laser ablation–gas chromatography–isotope ratio mass spectrometry following established methods (ref. 22, Supplementary Text, Supplementary Table 1, and Supplementary Figs 1 and 2). When necessary, sample surfaces were cleaned using abrasional removal of surface layers, treatment in 10% hydrogen peroxide or treatment in ethanol, or a combination of these. Data were normalized to CO₂ reference gas aliquots calibrated to NBS-19 (TS-Limestone). Accuracy and precision were also monitored through repeated analyses of an internal tooth enamel standard calibrated to NBS-19.

Microwear data were collected following established protocols (refs 15, 18, Supplementary Text and Supplementary Figs 3 and 4). Specimens (RM² of MH1 and lower right second molar (RM₂) of MH2) were cleaned with alcohol and cotton swabs, moulded using a polyvinylsiloxane impression material and cast in high-resolution epoxy. Point clouds with 0.18- μ m lateral spacing were generated using a white-light scanning confocal profiler representing four adjoining surfaces on facet 9, each measuring 102 \times 138 μ m. Scale-sensitive fractal complexity and length-scale anisotropy data were computed using Toothfrax software for each surface, and median values were calculated for each specimen.

Calculi from the buccal surfaces of the LI² and LP³ (upper left second incisor and upper left anterior premolar) of MH1 were collected using an established protocol (ref. 19, Supplementary Text and Supplementary Fig. 5) and analysed using transmitted light microscopy at $\times 400$ magnification. The samples were then processed further with hydrochloric acid to remove calcite crystals and then re-examined. Sediment from the breccia block encasing MH1 was examined to test for potential contamination (Supplementary Text). Sediment samples from the site were examined to provide an environmental baseline (ref. 19, Supplementary Text, Supplementary Table 4 and Supplementary Fig. 6). Phytoliths were identified based on comparison to modern plant reference collections.

Received 12 January; accepted 8 May 2012.

Published online 27 June 2012.

- Pickering, R. *et al.* *Australopithecus sediba* at 1.977 Ma and implications for the origins of the genus *Homo*. *Science* **333**, 1421–1423 (2011).
- Berger, L. R. *et al.* *Australopithecus sediba*: a new species of *Homo*-like australopithecine from South Africa. *Science* **328**, 195–204 (2010).
- Carlson, K. J. *et al.* The endocast of MH1, *Australopithecus sediba*. *Science* **333**, 1402–1407 (2011).
- Kibii, J. M. *et al.* A partial pelvis of *Australopithecus sediba*. *Science* **333**, 1407–1411 (2011).
- Kivell, T. L., Kibii, J. M., Churchill, S. E., Schmid, P. & Berger, L. R. *Australopithecus sediba* hand demonstrates mosaic evolution of locomotor and manipulative abilities. *Science* **333**, 1411–1417 (2011).
- Zipfel, B. *et al.* The foot and ankle of *Australopithecus sediba*. *Science* **333**, 1417–1420 (2011).
- Vogel, J. C. Isotopic assessment of the dietary habits of ungulates. *S. Afr. J. Sci.* **74**, 298–301 (1978).
- Schoeninger, M. J., Moore, J. & Sept, J. M. Subsistence strategies of two “savanna” chimpanzee populations: the stable isotope evidence. *Am. J. Primatol.* **49**, 297–314 (1999).
- Sponheimer, M. *et al.* Do “savanna” chimpanzees consume C_4 resources? *J. Hum. Evol.* **51**, 128–133 (2006).
- Lee-Thorp, J., Thackeray, F. & van der Merwe, N. J. The hunters and the hunted revisited. *J. Hum. Evol.* **39**, 565–576 (2000).
- van der Merwe, N. J., Thackeray, J. F., Lee-Thorp, J. A. & Luyt, J. The carbon isotope ecology and diet of *Australopithecus africanus* at Sterkfontein, South Africa. *J. Hum. Evol.* **44**, 581–597 (2003).
- van der Merwe, N. J., Masao, F. T. & Bamford, M. K. Isotopic evidence for contrasting diets of early hominins *Homo habilis* and *Australopithecus boisei* of Tanzania. *S. Afr. J. Sci.* **104**, 153–155 (2008).
- Ungar, P. S. *Mammal Teeth: Origin, Evolution, and Diversity* (Johns Hopkins Univ. Press, 2010).
- Scott, R. S. *et al.* Dental microwear texture analysis shows within-species diet variability in fossil hominins. *Nature* **436**, 693–695 (2005).
- Grine, F. E., Ungar, P. S., Teaford, M. F. & El-Zaatari, S. Molar microwear in *Præanthropus afarensis*: evidence for dietary stasis through time and under diverse paleoecological conditions. *J. Hum. Evol.* **51**, 297–319 (2006).
- Pontzer, H., Scott, R. S., Lordkipanidze, D. & Ungar, P. S. Dental microwear texture analysis and diet in the Dmanisi hominins. *J. Hum. Evol.* **61**, 683–687 (2011).
- Ungar, P. S., Scott, R. S., Grine, F. E. & Teaford, M. F. Molar microwear textures and the diets of *Australopithecus anamensis* and *Australopithecus afarensis*. *Phil. Trans. R. Soc. B* **365**, 3345–3354 (2010).
- Piperno, D. *Phytoliths: A Comprehensive Guide for Archaeologists and Paleoecologists* (AltaMira, 2006).
- Henry, A. G., Brooks, A. S. & Piperno, D. R. Microfossils in calculus demonstrate consumption of plants and cooked foods in Neanderthal diets (Shanidar III, Iraq; Spy I and II, Belgium). *Proc. Natl Acad. Sci. USA* **108**, 486–491 (2011).
- Bamford, M. *et al.* Botanical remains from a coprolite from the Pleistocene hominin site of Malapa, Sterkfontein Valley, South Africa. *Palaeontol. Afr.* **45**, 23–28 (2010).
- Dirks, P. H. *et al.* Geological setting and age of *Australopithecus sediba* from southern Africa. *Science* **328**, 205 (2010).
- Passey, B. H. & Cerling, T. E. In situ stable isotope analysis ($\delta^{13}C$, $\delta^{18}O$) of very small teeth using laser ablation GC/IRMS. *Chem. Geol.* **235**, 238–249 (2006).
- Lee-Thorp, J. A., van der Merwe, N. J. & Brain, C. Diet of *Australopithecus robustus* at Swartkrans from stable carbon isotopic analysis. *J. Hum. Evol.* **27**, 361–372 (1994).
- Sponheimer, M. *et al.* Hominins, sedges, and termites: New carbon isotope data from the Sterkfontein valley and Kruger National Park. *J. Hum. Evol.* **48**, 301–312 (2005).
- Cerling, T. E. *et al.* Diet of *Paranthropus boisei* in the early Pleistocene of East Africa. *Proc. Natl Acad. Sci. USA* **108**, 9337–9347 (2011).

26. White, T. D. *et al.* Macrovertebrate paleontology and the Pliocene habitat of *Ardipithecus ramidus*. *Science* **326**, 87–93 (2009).
27. Smith, C. C., Morgan, M. E. & Pilbeam, D. Isotopic ecology and dietary profiles of Liberian chimpanzees. *J. Hum. Evol.* **58**, 43–55 (2010).
28. Lambert, J. E., Chapman, C. A., Wrangham, R. W. & Conklin-Brittain, N. L. Hardness of cercopithecine foods: implications for the critical function of enamel thickness in exploiting fallback foods. *Am. J. Phys. Anthropol.* **125**, 363–368 (2004).
29. Rogers, M. E. *et al.* *Current Primatology Volume 1: Ecology and Evolution* (eds Thierry, B., Anderson, J. R., Roeder, J. J. & Herrenschmidt, N.) 37–43 (Universite Louis Pasteur, 1994).
30. Ungar, P. S. & Sponheimer, M. The diets of early hominins. *Science* **334**, 190–193 (2011).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements A.G.H. was funded in part by the US National Science Foundation (NSF), the Smithsonian Institution, the Malapa Project at the Institute for Human Evolution of the University of the Witwatersrand, and the Max Planck Society. P.S.U. was funded by the US Department of State Fulbright Scholarship Program, the Malapa Project and the US NSF. B.H.P. was supported by the US NSF. M.S. was funded by the US NSF, the Leakey Foundation and the University of Colorado Dean's Fund for Excellence. The phytolith reference material and processing of M.B. was

funded by the Palaeontological Scientific Trust, South Africa, and the South African Department of Science and Technology. P.S. was supported by the Wenner-Gren Foundation. D.J.d.R. was funded by the Ray A. Rothrock Fellowship and the International Research Travel Assistance Grant of Texas A&M University. Extensive funding for the Malapa Project comes from the Gauteng Provincial Government, the South African Department of Science and Technology, the South Africa National Research Foundation, the University of the Witwatersrand, and many private and public donors. We thank K. Krueger for running the microwear texture data through ToothFrax and SFrax, and J. Leichliter and O. Paine for their work with the Malapa roost rodents.

Author Contributions A.G.H., P.S.U., M.S., D.J.d.R. and L.B. conceived the project. A.G.H. collected and processed sediment and calculus samples. P.S.U. performed the dental microwear analysis. B.H.P., M.S. and P.S. performed isotopic analyses. L.R. identified grass phytoliths in the calculus and examined grass silica short cells (GSSCs) in Facies D sediment sample. M.B. identified the non-grass phytoliths in the calculus and analysed sediment samples. All authors contributed to the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to A.G.H. (amanda_henry@eva.mpg.de).

Compartmentalized calcium dynamics in a *C. elegans* interneuron encode head movement

Michael Hendricks¹, Heonick Ha¹, Nicolas Maffey¹ & Yun Zhang¹

The confinement of neuronal activity to specific subcellular regions is a mechanism for expanding the computational properties of neurons. Although the circuit organization underlying compartmentalized activity has been studied in several systems^{1–4}, its cellular basis is still unknown. Here we characterize compartmentalized activity in *Caenorhabditis elegans* RIA interneurons, which have multiple reciprocal connections to head motor neurons and receive input from sensory pathways. We show that RIA spatially encodes head movement on a subcellular scale through axonal compartmentalization. This subcellular axonal activity is dependent on acetylcholine release from head motor neurons and is simultaneously present and additive with glutamate-dependent globally synchronized activity evoked by sensory inputs. Postsynaptically, the muscarinic acetylcholine receptor GAR-3 acts in RIA to compartmentalize axonal activity through the mobilization of intracellular calcium stores. The compartmentalized activity functions independently of the synchronized activity to modulate locomotory behaviour.

The ‘wiring diagram’ of the *C. elegans* nervous system allows the dissection of circuit function from single neurons to behavioural outputs^{5,6}. RIA interneurons regulate navigation behaviours and occupy a key circuit position, receiving input from multiple sensory

networks and showing reciprocal synaptic connections with head motor neurons (Fig. 1a)^{5,7,8}. To characterize the properties of RIA, we expressed the genetically encoded calcium indicator GCaMP3 (ref. 9) under the RIA-specific *glr-3* promoter¹⁰ and performed calcium imaging with a microfluidic device in which semi-restrained transgenic animals could bend their heads in the dorsal–ventral plane, allowing simultaneous monitoring of neural activity and head movements¹¹. We stimulated the animals with alternating fluid streams of 3-methylbutan-1-ol (also known as isoamyl alcohol; IAA) and buffer, because IAA is detected by AWC olfactory neurons, one of the sensory neurons upstream of RIA and some of its major synaptic inputs¹². We did not detect any prominent calcium response in the RIA cell body (Supplementary Fig. 1a). The single axon of RIA projects ventrally into the nerve cord, where it forms a hairpin loop, then enters and extends around the nerve ring, forming a near-complete circle (Fig. 1b). *C. elegans* normally lies on its side such that several distinct portions of the RIA axon lie in the same focal plane: the ‘loop’ (the ventrally directed hairpin loop), nrD and nrV (the dorsal and ventral segments of the RIA axon in the nerve ring, respectively) (Fig. 1b). Strikingly, we observed robust calcium dynamics in each of these axonal segments that seemed to be independent, and this axonal calcium activity was correlated with head movement (Fig. 1c, d and Supplementary Movie 1).

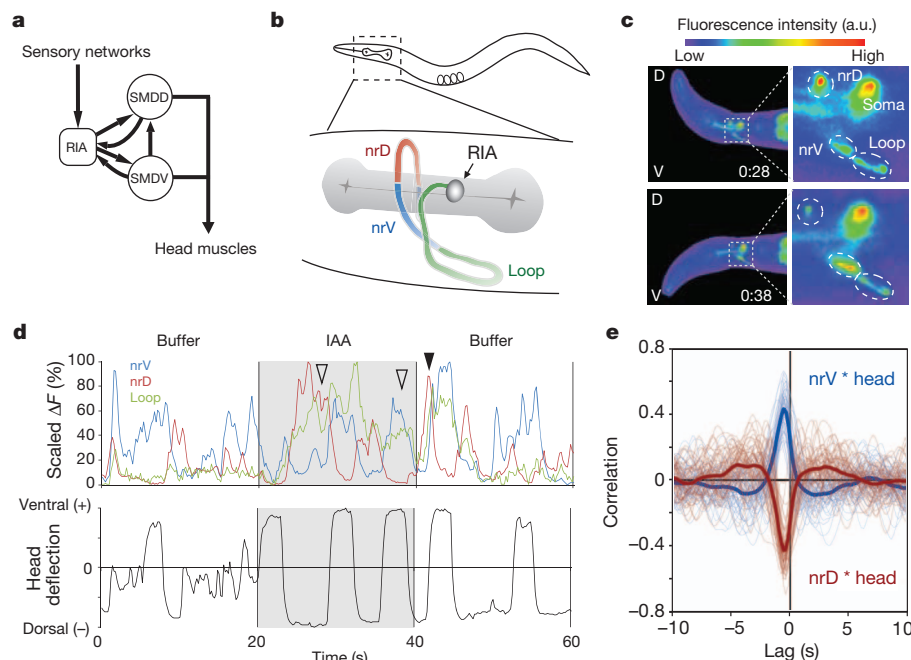


Figure 1 | Compartmentalized calcium signals in RIA encode head movement. **a**, Highly simplified RIA and SMD synaptic connections. **b**, RIA anatomy showing nrD, nrV and loop subdomains. **c**, Single frames of GCaMP3 fluorescence recording in RIA; times are shown in the format mins:s. Dashed ovals denote subdomains in **b**. Anterior is to the left. D, dorsal; V, ventral.

d, Calcium dynamics in RIA axonal domains and corresponding head bending. Open arrowheads, time points in **c**; filled arrowhead, a synchronous calcium event. **e**, Cross-correlations between nrV or nrD Ca^{2+} responses and head movement. Faint lines are results for individual animals ($n = 53$); the thick lines show mean values.

¹Department of Organismic and Evolutionary Biology, Center for Brain Science, Harvard University, Cambridge, Massachusetts 02138, USA.

Calcium transients in nrD were correlated with dorsal head bends, whereas calcium increase in nrV was correlated with ventral head bends. Although another interneuron, AIY, also shows calcium signals in its axon, but not its soma^{12,13}, under these conditions the calcium dynamics are synchronized along the entire process (Supplementary Fig. 1b–d).

To quantify the correlation between head movement and RIA calcium dynamics, we defined ventral head deflection as positive and dorsal deflection as negative and performed a cross-correlation analysis between head bending and calcium signals. We found a clear positive correlation between calcium activity in nrV and head movement, whereas calcium dynamics in nrD were negatively correlated with head movement (Fig. 1e). Consistently, calcium responses in nrD and nrV were uncorrelated (Supplementary Fig. 1e). The correlations between nrD and nrV calcium activity and head movement were independent of stimulus, and the correlations were not due to a shared dependence on the stimulus (Supplementary Fig. 1f, g). We also observed distinct calcium transients that occurred synchronously in all three RIA axonal domains (Fig. 1d, filled arrowhead), as discussed below. We recorded RIA axonal activity during forward crawling under a coverslip (Supplementary Fig. 2a) and measured GCaMP3 intensity throughout the RIA axon during head bends (Supplementary Fig. 2b). The results consistently showed strong correlations between calcium responses in nrV and nrD with ventral and dorsal head bending, respectively (Supplementary Fig. 2c). Thus, the single RIA axon is partitioned into distinct domains, and calcium dynamics within these domains spatially represent, or encode, head movement.

To investigate how circuit properties regulate compartmentalized activity in RIA, we examined the subcellular organization of the synapses of RIA within the nerve ring. Most of these synapses are reciprocal connections with the SMDV and SMDD motor neurons, which innervate ventral and dorsal head muscles, respectively⁵. SMDs also have electrical and chemical synapses with several other head motor neurons that innervate the same muscles and are involved in locomotion^{5,7,8}. SMD–RIA synapses are spatially segregated: SMDV synapses onto RIA only in nrV, whereas SMDD is presynaptic to RIA only in nrD (Fig. 2a). We expressed GCaMP3 in the SMDs by using the

glr-1 promoter¹⁴ and recorded calcium activity and head movement. As expected, the calcium responses in SMDV and SMDD were correlated with ventral and dorsal head bending, respectively (Fig. 2b, c and Supplementary Movie 2). Thus, compartmentalized calcium transients in RIA may be driven by SMDs. We also examined calcium responses in animals treated with the nicotinic acetylcholine receptor agonist levamisole. Robust compartmentalized dynamics in RIA and oscillatory dynamics in SMDs persisted in completely paralysed animals (Fig. 2d and Supplementary Movies 3 and 4). Thus, RIA axonal activity is independent of movement.

Next, we blocked neurotransmission from SMDD and SMDV neurons by expressing tetanus toxin (TeTx) from either the *glr-1* or *lad-2* promoter and examined the effects on RIA activity^{15,16}. Both the *glr-1* and *lad-2* promoters drive expression in multiple neurons, but among the presynaptic partners of RIA their expression overlaps only in the SMDs^{14,17}. Animals expressing either the *glr-1::TeTx* or *lad-2::TeTx* transgene were severely impaired in both nrV and nrD correlations with head movement without losing calcium dynamics, producing synchronized calcium dynamics throughout the RIA neuronal process (Fig. 2e–g). Intracellular cross-correlations between nrV and nrD calcium signals were high in these animals (Fig. 2f, g). TeTx expression in SMDs did not abolish head bending, probably as a result of functional redundancy with several other motor neurons. Thus, the motor-correlated compartmentalized RIA axonal activity is driven by SMD motor neurons.

Because SMDs are cholinergic, we examined the effect of a hypomorphic mutation, *cha-1(p1152)*, in choline acetyltransferase, which is required for the biosynthesis of acetylcholine and is expressed in all cholinergic motor neurons¹⁸. The *cha-1* mutants showed active calcium responses in RIA but were impaired in the correlations between RIA nerve ring calcium dynamics and head bending and defective in the compartmentalization of RIA axon (Fig. 3a, b, e). These defects were rescued by a cosmid transgene containing the *cha-1* genomic locus (Supplementary Fig. 2d). Thus, compartmentalized activity in nrV and nrD and its correlation with head movement require cholinergic neurotransmission.

Most *C. elegans* neurons are predicted to be isopotential¹⁹, making the involvement of postsynaptic ionotropic receptors in generating

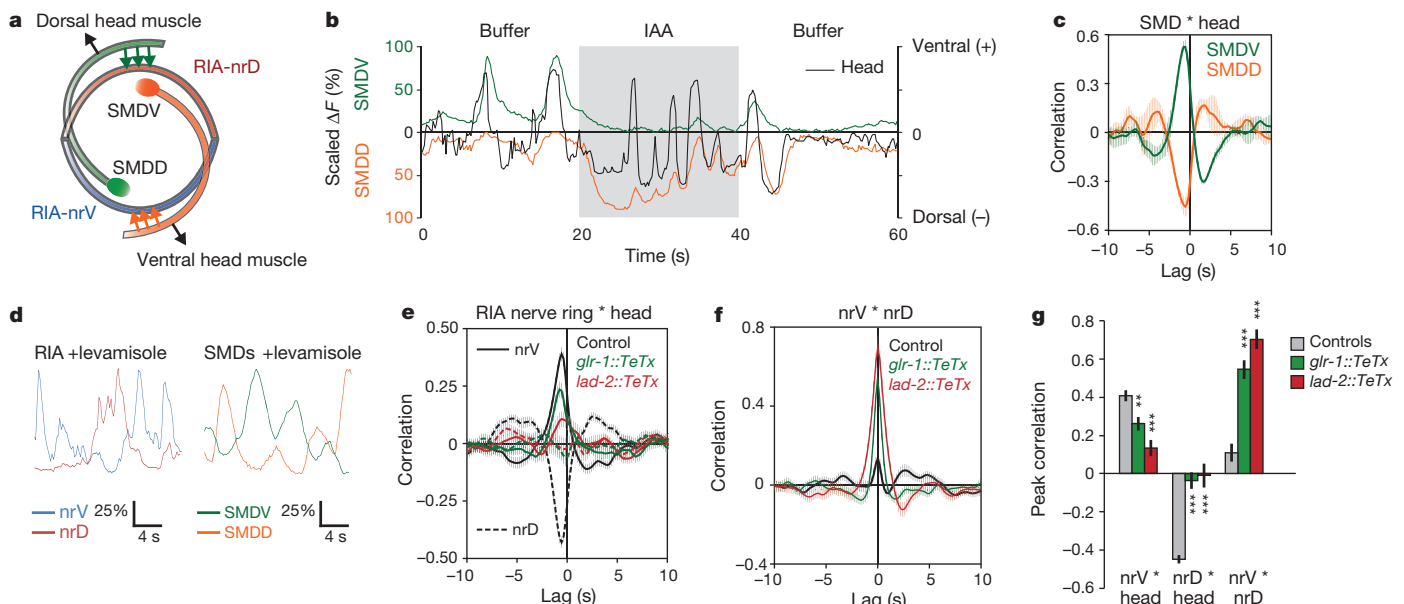


Figure 2 | RIA compartmentalized activity requires transmission from SMD motor neurons. **a**, Topography of SMD synapses onto RIA and muscles. **b**, Overlay of representative SMD calcium recording and movement. The SMDD trace is inverted to show correlation with dorsal head bending. **c**, Cross-correlations of SMDV and SMDD with head movement ($n = 8$). **d**, Representative calcium recordings of RIA subdomains or SMDs in

levamisole-treated animals. **e**, RIA correlations with head movement in *glr-1::TeTx* or *lad-2::TeTx* transgenic strains (*glr-1*, $n = 18$; *lad-2*, $n = 17$; control, $n = 28$). **f**, Internal cross-correlations in SMD::TeTx strains. **g**, Comparison of peak correlations between strains. Cross-correlation plots and bar charts are means \pm s.e.m. Three asterisks, $P < 0.001$; two asterisks, $P < 0.01$ (analysis of variance corrected for multiple comparisons).

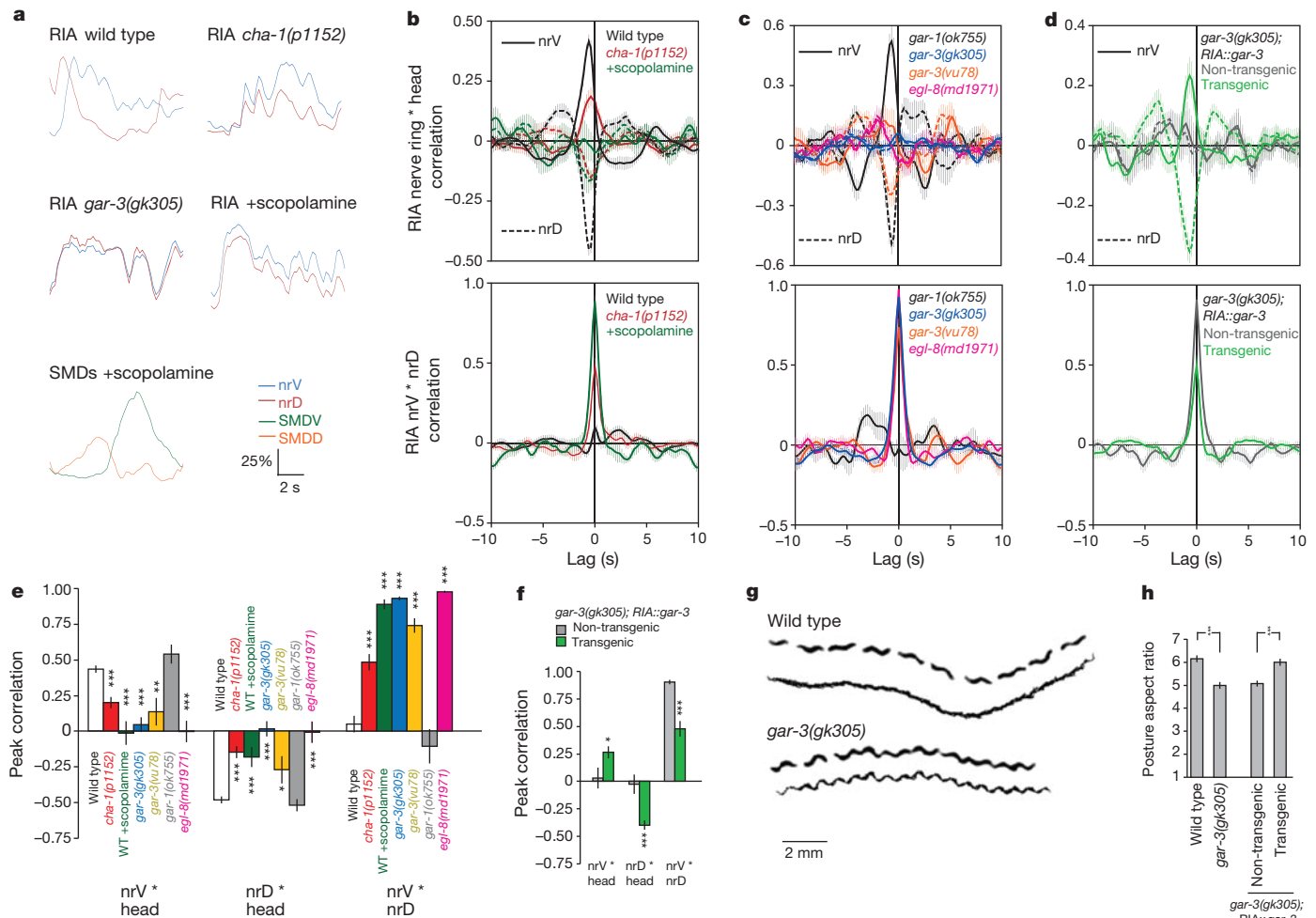


Figure 3 | A muscarinic acetylcholine pathway establishes motor-correlated RIA dynamics. **a**, Sample calcium dynamics of nrV, nrD and SMDs. **b**, Cross-correlations between RIA dynamics and head movement and internal cross-correlations in *cha-1(p1152)* ($n = 18$; control, $n = 25$) and scopolamine-treated wild-type animals ($100 \mu\text{M}$, $n = 10$). **c**, Cross-correlation plots as in **b** for mutants: *gar-1(ok755)* ($n = 5$), *gar-3(gk305)* ($n = 16$), *gar-3(vu78)* ($n = 6$) and *egl-8(md1971)* ($n = 7$). **d**, Cross-correlations of transgenic ($n = 13$) and non-transgenic ($n = 9$) siblings in a strain expressing a *gar-3* cDNA in RIA in *gar-3(gk305)* mutants. **e**, Peak correlation comparisons for data in **b** and **c**.

f, Peak correlation comparisons for data in **d**. WT, wild type. **g**, Images of wild-type and *gar-3(gk305)* animals every 5 s over 50 s; beneath each image series is the projection of each animal's midline every 0.5 s over the same period. **h**, Comparison of mean postural aspect ratios during forward locomotion for wild-type ($n = 17$), *gar-3(gk305)* ($n = 10$), transgenic ($n = 10$) and non-transgenic ($n = 15$) siblings in the RIA::*gar-3* rescue line. Cross-correlation plots and bar charts are means \pm s.e.m. Three asterisks, $P < 0.001$; two asterisks, $P < 0.01$; asterisk, $P < 0.05$ (analysis of variance corrected for multiple comparisons).

local calcium changes unlikely. *C. elegans* has a single class of G-protein-coupled muscarinic acetylcholine receptors (mAChRs), namely GAR-1, GAR-2 and GAR-3 (ref. 20). When wild-type animals were briefly treated with the GAR-3-selective antagonist scopolamine²⁰, we observed a loss of compartmentalization and head movement correlations and a strong correlation between nrV and nrD dynamics (Fig. 3a, b, e). Treatment with scopolamine had no effect on SMD dynamics or their correlation with head movement, demonstrating that the loss of RIA compartmentalization is not due to impaired SMD function (Fig. 3a and Supplementary Fig. 2e). Consistently, the correlation between head movement and RIA axonal activities was abolished in the loss-of-function mutant *gar-3(gk305)*²¹ and significantly decreased in *gar-3(vu78)*, which harbours a missense mutation²². Both mutants displayed strong correlations between nrV and nrD activity (Fig. 3a, c, e). In contrast, *gar-1(ok755)* mutants did not show defects in RIA calcium responses (Fig. 3c, e). RIA-specific expression of a wild-type *gar-3* complementary DNA in *gar-3(gk305)* mutants significantly rescued these defects (Fig. 3d, f), demonstrating the postsynaptic function of GAR-3 in generating compartmentalization of RIA axonal activity. GAR-3 is similar to the mammalian M1/3/5 mAChRs in both its pharmacology and its downstream signalling

mechanism, which is mediated by $G\alpha_q$ and phospholipase C β (PLC- β)^{20,23}. There is only one neuronal PLC- β orthologue, EGL-8, in *C. elegans*^{24,25}. Consistently, the nonsense mutation in *egl-8(md1971)*²⁴ showed a defect in RIA compartmentalization similar to that in *gar-3(gk305)* mutants (Fig. 3c, e). PLC- β is known to regulate neural activity by activating calcium stores in the endoplasmic reticulum. Taken together, our results are consistent with muscarinic signalling through GAR-3 in RIA mobilizing internal calcium stores via PLC- β to establish compartmentalized axonal activity.

The *gar-3(gk305)* mutant allowed us to selectively test the behavioural consequences of RIA compartmentalized activity, because it does not disrupt synchronous responses or grossly disturb RIA-associated circuitry. We recorded animals during forward locomotion and found that *gar-3(gk305)* mutants generated deeper head bending, leading to a larger-amplitude body wave (Fig. 3g). We calculated the mean aspect ratio of an ellipse fitted to the animals' posture in each movie frame. The *gar-3(gk305)* mutants showed a significantly lower posture aspect ratio than wild-type animals, and this defect was fully rescued by expression of a wild-type *gar-3* cDNA in RIA (Fig. 3h). Thus, the compartmentalized axonal activity of RIA regulates head movement.

In addition, we observed short, synchronous calcium events across all three RIA axonal compartments (Fig. 1d). We examined the temporal distribution of these events by defining synchronous calcium influx and efflux events as time points at which the rate of change in GCaMP3 intensity in all three compartments was positive and negative, respectively. We subjected animals to alternating streams of buffer and IAA and scored synchronous events in the raster plot shown in Fig. 4a. We calculated the mean rate of calcium flux across all three axonal domains (time histogram in Fig. 4a). Two trials per animal were averaged to generate the time histograms shown in Fig. 4b. Removal of IAA (IAA-off) evoked synchronous calcium influx followed by synchronous efflux (filled arrowhead and open arrow, respectively, in Fig. 4a). In contrast, switching from buffer to IAA (IAA-on) elicited a high frequency of synchronous calcium efflux events followed by

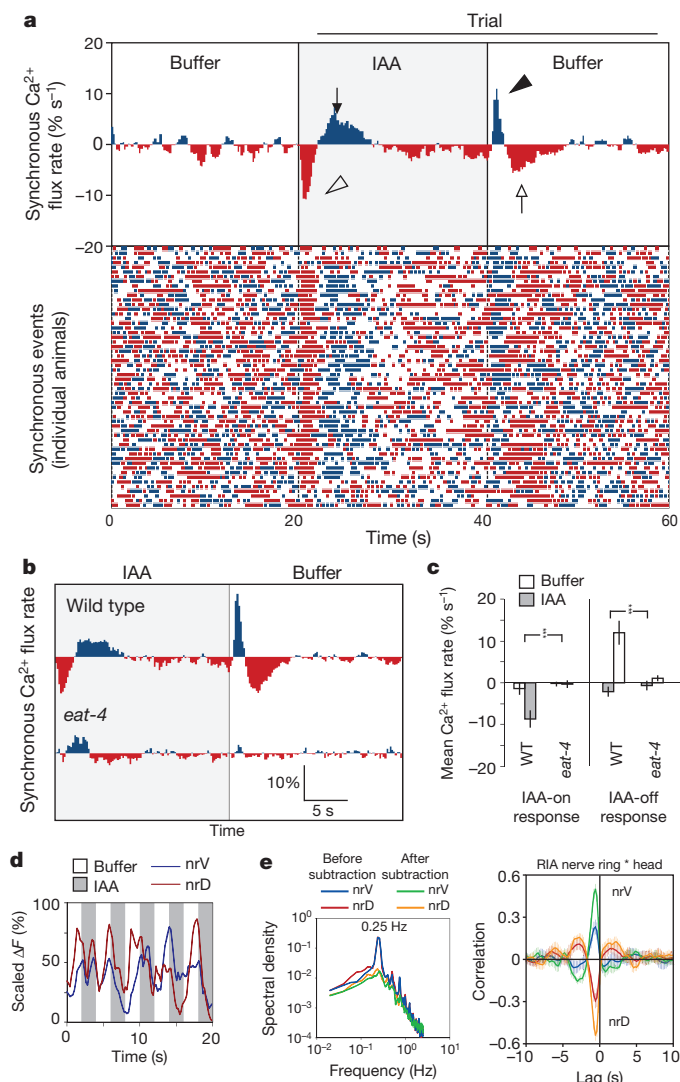


Figure 4 | Sensory-evoked synchronized RIA activity is dependent on glutamate-dependent transmission and is additive with compartmentalized dynamics. **a**, Raster plot and histogram of synchronous calcium influx (blue) or efflux (red). Dots show synchronous events; the histogram shows the mean rate of Ca²⁺ flux ($n = 53$). **b**, **c**, Histogram (**b**) and bar charts (**c**) for mean responses across two trials in wild-type (WT; $n = 19$) and *eat-4*(*ky5*) mutants ($n = 18$). Three asterisks, $P < 0.001$ for significant interaction between stimulus response and genotype (multiple analysis of variance with repeated measures). Error bars represent s.e.m. **d**, Example of nrV and nrD responses to switching between IAA and buffer every 2 s. **e**, Spectral density of nrV and nrD activity (left) and cross-correlations (right; means \pm s.e.m.) between nrV or nrD and head movement before and after synchronous signal subtraction (see the text).

influx (open arrowhead and filled arrow, respectively, in Fig. 4a). Thus, global RIA axonal activity is transiently suppressed by IAA presentation and activated by its removal.

IAA is sensed by the AWC olfactory sensory neurons, and sensorimotor responses to IAA require glutamate-dependent neurotransmission¹². Mutants defective in the vesicular glutamate transporter EAT-4 showed a complete loss of IAA-evoked activity in RIA (Fig. 4b, c); this was rescued by an *eat-4* transgene (Supplementary Fig. 3a). Motor-correlated compartmentalized RIA axonal dynamics were maintained in the *eat-4* mutants (Supplementary Fig. 3b). Blocking synaptic release from SMDs did not disrupt synchronous responses (Supplementary Fig. 4). Thus, sensory-evoked synchronous activity in RIA is regulated independently of the motor neuron inputs into nrV and nrD compartments. The synchronous and compartmentalized calcium dynamics in RIA may be either simultaneous and additive or mutually exclusive. To distinguish between these possibilities, we challenged RIA with high-frequency alternating stimuli (0.25 Hz) to see whether head movement correlations were maintained against rapid synchronous activity. Fast switching caused synchronous calcium dynamics in phase with the stimuli, the spectral profiles of nrV and nrD calcium signals showed 0.25-Hz peaks corresponding to the stimulus, and correlations between nrV and nrD and head movement seemed reduced (Fig. 4d, e). After subtraction of the synchronous component from each compartment, the correlations between head movement and nrV and nrD activity were restored (Fig. 4e), demonstrating that the calcium signals measured in each RIA axonal compartment represented the combined signals of sensory-evoked and motor inputs.

There are two primary methods for the nervous system to monitor its own output: proprioceptive feedback, which responds to the mechanical stimuli produced by self-motion, and corollary discharge, in which a copy of the motor command is transmitted to sensory processing networks²⁶. RIA compartmentalized dynamics possess key features of corollary discharge: they originate from the motor system, represent spatial attributes of the behaviour, and modulate the encoded behaviour. Several *C. elegans* behaviours require the detection of spatial changes in stimuli relative to head movement. Network models for these behaviours have been proposed; however, the potential underlying circuit mechanisms have not been observed *in vivo*²⁷. Our identification of corollary discharge from head motor neurons to interneurons adds an important property to the navigation network, by which a sensorimotor circuit can monitor self-motion by means of corollary discharge and integrate motor dynamics with sensory stimuli to regulate locomotion (Supplementary Fig. 5).

Compartmentalization is a feature of morphologically complex cells, such as vertebrate starburst amacrine cells⁴ and fly tangential cells²⁸. The existence of compartmentalization within the unbranched axon of a unipolar neuron in *C. elegans* suggests a deeply conserved mechanism for signal processing. Our identification of a cholinergic pathway establishing RIA compartmental activity agrees with observations in inhibitory thalamic interneurons, in which cholinergic input regulates the dynamic coupling between distal dendrites and the rest of the neuron⁴. These comparisons suggest commonalities in the signalling and circuit properties underlying compartmentalization. Our findings provide a framework for sensorimotor processing and integration at the subcellular level and a mechanism underlying functional specificity of circuit connections among *C. elegans* neurons, many of which are isopotential, thus revealing a new perspective on the network topology of the nematode nervous system.

METHODS SUMMARY

Strains were maintained under standard conditions²⁹. Transgenesis³⁰ and calcium imaging^{7,11,12} were performed as described. Cross-correlations and statistical analyses were performed with JMP8 (SAS). Time series data were filtered by exponential smoothing, an instantaneous time derivative was calculated for each frame, and these data were used for cross-correlation. To detect synchronous calcium

peaks, we scored synchronous events and calculated the mean rate of calcium flux as described in the text. Locomotion analysis was performed on NGM agar without food, using a 5-megapixel camera at 2 frames per second.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 16 May 2011; accepted 22 March 2012.

Published online 13 May 2012.

- Euler, T., Detwiler, P. B. & Denk, W. Directionally selective calcium signals in dendrites of starburst amacrine cells. *Nature* **418**, 845–852 (2002).
- Borst, A. & Euler, T. Seeing things in motion: models, circuits, and mechanisms. *Neuron* **71**, 974–994 (2011).
- Rall, W., Shepherd, G. M., Reese, T. S. & Brightman, M. W. Dendrodendritic synaptic pathway for inhibition in the olfactory bulb. *Exp. Neurol.* **14**, 44–56 (1966).
- Zhu, J. & Heggelund, P. Muscarinic regulation of dendritic and axonal outputs of rat thalamic interneurons: a new cellular mechanism for uncoupling distal dendrites. *J. Neurosci.* **21**, 1148–1159 (2001).
- White, J. G., Southgate, E., Thomson, J. N. & Brenner, S. The structure of the nervous system of the nematode *Caenorhabditis elegans*. *Phil. Trans. R. Soc. Lond. B* **314**, 1–340 (1986).
- Varshney, L. R., Chen, B. L., Paniagua, E., Hall, D. H. & Chklovskii, D. B. Structural properties of the *Caenorhabditis elegans* neuronal network. *PLOS Comput. Biol.* **7**, e1001066 (2011).
- Ha, H.-I. *et al.* Functional organization of a neural network for aversive olfactory learning in *Caenorhabditis elegans*. *Neuron* **68**, 1173–1186 (2010).
- Gray, J. M., Hill, J. J. & Bargmann, C. I. A circuit for navigation in *Caenorhabditis elegans*. *Proc. Natl Acad. Sci. USA* **102**, 3184–3191 (2005).
- Tian, L. *et al.* Imaging neural activity in worms, flies and mice with improved GCaMP calcium indicators. *Nature Methods* **6**, 875–881 (2009).
- Brockie, P. J., Madsen, D. M., Zheng, Y., Mellem, J. & Maricq, A. V. Differential expression of glutamate receptor subunits in the nervous system of *Caenorhabditis elegans* and their regulation by the homeodomain protein UNC-42. *J. Neurosci.* **21**, 1510–1522 (2001).
- Chronis, N., Zimmer, M. & Bargmann, C. I. Microfluidics for *in vivo* imaging of neuronal and behavioral activity in *Caenorhabditis elegans*. *Nature Methods* **4**, 727–731 (2007).
- Chalasani, S. H. *et al.* Dissecting a circuit for olfactory behaviour in *Caenorhabditis elegans*. *Nature* **450**, 63–70 (2007).
- Clark, D. A., Biron, D., Sengupta, P. & Samuel, A. D. T. The AFD sensory neurons encode multiple functions underlying thermotactic behavior in *Caenorhabditis elegans*. *J. Neurosci.* **26**, 7444–7451 (2006).
- Hart, A. C., Sims, S. & Kaplan, J. M. Synaptic code for sensory modalities revealed by *C. elegans* GLR-1 glutamate receptor. *Nature* **378**, 82–85 (1995).
- Macosko, E. Z. *et al.* A hub-and-spoke circuit drives pheromone attraction and social behaviour in *C. elegans*. *Nature* **458**, 1171–1175 (2009).
- Schiavo, G. *et al.* Tetanus and botulinum-B neurotoxins block neurotransmitter release by proteolytic cleavage of synaptobrevin. *Nature* **359**, 832–835 (1992).
- Wang, X. *et al.* The *C. elegans* L1CAM homologue LAD-2 functions as a coreceptor in MAB-20/Sema2 mediated axon guidance. *J. Cell Biol.* **180**, 233–246 (2008).
- Rand, J. B. & Nonet, M. L. in *C. elegans II* (eds Riddle, D. L., Blumenthal, T., Meyer, B. J. & Priess, J. R.) 611–643 (Cold Spring Harbor Laboratory Press, 1997).
- Goodman, M. B., Hall, D. H., Avery, L. & Lockery, S. R. Active currents regulate sensitivity and dynamic range in *C. elegans* neurons. *Neuron* **20**, 763–772 (1998).
- Park, Y.-S., Kim, S., Shin, Y., Choi, B. & Cho, N. J. Alternative splicing of the muscarinic acetylcholine receptor GAR-3 in *Caenorhabditis elegans*. *Biochem. Biophys. Res. Commun.* **308**, 961–965 (2003).
- Liu, Y., LeBoeuf, B. & Garcia, L. R. G_{α_q} -coupled muscarinic acetylcholine receptors enhance nicotinic acetylcholine receptor signaling in *Caenorhabditis elegans* mating behavior. *J. Neurosci.* **27**, 1411–1421 (2007).
- Steger, K. A. & Avery, L. The GAR-3 muscarinic receptor cooperates with calcium signals to regulate muscle contraction in the *Caenorhabditis elegans* pharynx. *Genetics* **167**, 633–643 (2004).
- Hulme, E. C., Birdsall, N. J. & Buckley, N. J. Muscarinic receptor subtypes. *Annu. Rev. Pharmacol. Toxicol.* **30**, 633–673 (1990).
- Miller, K. G., Emerson, M. D. & Rand, J. B. G_{α_q} and diacylglycerol kinase negatively regulate the G_{α_q} pathway in *C. elegans*. *Neuron* **24**, 323–333 (1999).
- Lackner, M. R., Nurris, S. J. & Kaplan, J. M. Facilitation of synaptic transmission by EGL-30 G_{α_q} and EGL-8 PLC β : DAG binding to UNC-13 is required to stimulate acetylcholine release. *Neuron* **24**, 335–346 (1999).
- Crapse, T. B. & Sommer, M. A. Corollary discharge across the animal kingdom. *Nature Rev. Neurosci.* **9**, 587–600 (2008).
- Lockery, S. R. The computational worm: spatial orientation and its neuronal basis in *C. elegans*. *Curr. Opin. Neurobiol.* **21**, 782–790 (2011).
- Joesch, M., Schnell, B., Raghu, S. V., Reiff, D. F. & Borst, A. ON and OFF pathways in *Drosophila* motion vision. *Nature* **468**, 300–304 (2010).
- Brenner, S. The genetics of *Caenorhabditis elegans*. *Genetics* **77**, 71–94 (1974).
- Mello, C. C., Kramer, J. M., Stinchcomb, D. & Ambros, V. Efficient gene transfer in *C. elegans*: extrachromosomal maintenance and integration of transforming sequences. *EMBO J.* **10**, 3959–3970 (1991).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the *Caenorhabditis* Genetics Center and *C. elegans* Gene Knockout Consortium for *C. elegans* strains; the Wellcome Trust Sanger Institute for cosmid; C. Bargmann for the *TeTx* cDNA, *eat-4* cDNA and *Pttx-3::GCaMP* strain; Y. Kohara for the *gar-3* cDNA; J. Nakai for the GCaMP DNA; N. Bhatla for help with integration; L. Tian and L. Looger for the GCaMP3.3 plasmid; Y. Shen for the *lad-2* promoter; and C. Bargmann, K. Blum, F. Engert, C. Fang-Yen, S. Hendricks, S. Jesuthasan, A. Samuel, J. Sanes, E. Soucy and members of the Zhang laboratory for advice and discussion. M.H. thanks S. Moser for personal support. This work was supported by funding from The Esther A. and Joseph Klingenstein Fund, the March of Dimes Foundation, the Alfred P. Sloan Foundation, the John Merck Fund and the National Institutes of Health (DC009852) to Y.Z.

Author Contributions M.H. and Y.Z. designed experiments, analysed data, interpreted results and wrote the paper. M.H., H.H. and N.M. performed experiments.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to Y.Z. (yzhang@oeb.harvard.edu).

METHODS

Strains. *Caenorhabditis elegans* strains were maintained under standard conditions²⁹. The following strains were used in this study: N2 (Bristol), ZC1507 *yxIs18* [*Pglr-3a::GCaMP3.3*, *Punc-122::dsred*], ZC1508 *yxIs19* [*Pglr-3a::GCaMP3.3*, *Punc-122::dsred*], ZC1148 *yxIs1* [*Pglr-1::GCaMP3.3*, *Punc-122::gfp*], ZC1591 *yxIs19*; *yxEx778* [*Plad-2::TeTx:mCherry*, *Punc-122::gfp*], MT6308 *eat-4(ky5)III*, PR1152 *cha-1(p1152)IV*, ZC1576 *cha-1(p1152)IV*; *yxIs19*, ZC1598 *eat-4(ky5)III*; *yxIs19*, ZC1563 *yxIs19*; *yxEx749* [*Pglr-1::TeTx:mCherry*, *Punc-122::gfp*], ZC1626 *eat-4(ky5)III*; *yxIs19*; *yxEx802* [*Peat-4::eat-4*, *Punc-122::gfp*], ZC1633 *cha-1(p1152)IV*; *yxIs19*; *yxEx809* [*cosmid ZC416*, *Punc-122::gfp*], RM2221 *egl-8(md1971)V*, ZC1922 *egl-8(md1971)V*; *yxIs19*, VC657 *gar-3(gk305)V*, JD217 *gar-3(vu78)V*, ZC1790 *gar-1(ok755)X*; *yxIs19*, ZC1792 *gar-3(gk305)V*; *yxIs19*, ZC1834 *gar-3(vu78)V*; *yxIs19*, ZC1923 *gar-3(gk305)V*; *yxIs19*; *yxEx968* [*Pglr-3::gar-3*].

Molecular biology. The genetically encoded calcium indicator *GCaMP3.3* (ref. 9) (a gift from L. Looger and L. Tian) was cloned upstream of the 3' untranslated region (UTR) of the *unc-54* gene in pSM_Not (a gift from C. Bargmann) and a Gateway recombination cassette (rFB) (Invitrogen) was ligated upstream of *GCaMP3.3* to produce pSM-rFB-*GCaMP3.3*. The *GCaMP3.3* sequence was replaced with a tetanus toxin-mCherry fusion (a gift from C. Bargmann) to produce pSM-rFB-*TeTx:mCherry*, or with a full-length *gar-3* cDNA (a gift from Y. Kohara) to produce pSM-rFB-*gar-3*. The rFB cassette was also cloned upstream of *eat-4* cDNA (a gift from C. Bargmann) to produce pSM-rFB-*eat-4*. A 5.3-kilobase (kb) genomic fragment upstream of the *glr-1* gene (forward primer 5'-ACTGACCACTGCAGC ATTTT-3', reverse 5'-TGTGAATGTGTCAGATTGGG-3'), 2.8 kb upstream of *glr-3* (forward 5'-TCGGAAATGCGGAAGTTCTT-3', reverse 5'-ATGTTAATA GCAAATATTGAAG-3') and 5.5 kb upstream of *eat-4* (forward 5'-TTGTCCA CATTCTAGTAG-3', reverse 5'-GGTTTCTGAAAATGATGATG-3') were amplified by polymerase chain reaction and cloned into the pCR8 Gateway entry vector (Invitrogen). Recombination reactions were performed in accordance with the manufacturer's instructions, producing *Pglr-3::GCaMP3.3*, *Pglr-1::GCaMP3.3*, *Pglr-1::TeTx:mCherry*, *Plad-2::TeTx:mCherry* and *Peat-4::eat-4*. A PCR product comprising 1.2 kb of the *glr-3* upstream sequence (*Pglr-3a*), *GCaMP3.3* and the *unc-54* 3' UTR (forward 5'-GCCAGAGATGAGCATGATTC-3', reverse 5'-CGGCCGACTAGTAGGAAACA-3') was used at 50 ng μl^{-1} for germline transformation. For *cha-1* genomic rescue, cosmid ZC416 (Wellcome Trust Sanger Institute) was injected at 10 ng μl^{-1} . Other plasmids were injected at 25 ng μl^{-1} . Either *Punc-122::gfp* or *Punc-122::dsred* was used as a co-injection marker for all lines. Microinjection was performed as described³⁰. Integration of extrachromosomal arrays was performed by γ -irradiation as described by the Hobert laboratory (http://biochemistry.hs.columbia.edu/labs/hobert/integration_protocol.html). Mutant lines were provided by the *Caenorhabditis* Genetics Center. For RIA imaging in mutant backgrounds, mutants were crossed to strain ZC1508, and F₂ progeny were selected for homozygosity for both the integrated transgene and mutation of interest.

Calcium imaging. Calcium imaging was performed in a microfluidic device essentially as described^{11,12}. Fluorescence time-lapse imaging (100-ms exposures, 5 Hz) was performed on a Nikon Eclipse Ti-U inverted microscope with a 40 \times oil-immersion objective and a Photometrics CoolSnap EZ camera. Animals were presented with alternating streams of NGM buffer (1 mM CaCl₂, 1 mM MgSO₄, 25 mM KPO₄ pH 6.0) and 3-methylbutan-1-ol (10⁻⁴ v/v) in NGM buffer, a concentration shown to be optimal for activating AWC¹². All image analysis was performed with NIH ImageJ, and frames were aligned using the StackReg plugin. After background

subtraction, total fluorescence intensity was measured from individual regions of interest (ROIs) corresponding to subcellular domains in the case of RIA or the entire cell body for other neurons. For RIA, intensity values for each ROI were linearly scaled from 0 to 1 with the formula $(F - F_{\min}) / (F_{\max} - F_{\min})$. Freely moving animals on *Escherichia coli* OP50 lawns were lightly compressed with a coverslip and imaged under a 10 \times objective with 10-ms exposures at 5 Hz. Frames were chosen in which the animal's head was deflected dorsally or ventrally. A linear ROI tracing the RIA process was used to generate an intensity profile. Where noted, animals were treated with 2 mM levamisole (Tetramisole HCl; Sigma L9756) or 100 μM scopolamine (Sigma S1875) immediately before and during imaging.

Quantification of head movement. An ROI comprising the mobile portion of the animal's head and centred on its midline was selected and binarized, whereby all pixels corresponding to the animal's head were converted to 1 and all background pixels to 0. A macro in ImageJ was used to calculate head displacement from the ROI centre in the dorsal-ventral plane normalized relative to the maximum deflection (in either direction). In brief, a centre of mass was defined as the mean pixel position along the dorsoventral axis. The difference between the centre of mass and the position of the midline was calculated for each frame. This gave an index between -1 and 1; ventral was defined as positive.

Correlation analysis. Cross-correlations and all statistical analyses were performed with JMP8 software (SAS). Fluorescence intensity and head position time series data were filtered by exponential smoothing ($\alpha = 0.3$), an instantaneous time derivative calculated for each frame, and these data were used for cross-correlation with 100 time lags (20 s). For comparison of peak correlations, the maximum correlation (positive or negative) in a 1-s time window centred on the peak of the control mean correlation was used. Mutants and transgenic animals were compared by analysis of variance with wild-type or non-transgenic siblings.

Analysis of synchronous calcium events. To detect short-term synchronous calcium peaks, we scored each time point in which the nrV, nrD and loop regions all had time derivatives greater than 0 (influx) or less than 0 (efflux) as a synchronous event (raster plot in Fig. 4a). At each of these time points, the mean rate of calcium flux (positive or negative) was calculated; time points without synchronous activity were scored 0. The mean rate of change across all animals was calculated at each time point. For each animal, a set of two IAA-on events and two IAA-off events were scored by comparing a 1-s time window just before the stimulus switch with a 1-s time window centred on the time point showing peak response in wild-type animals. Analysis of variance with repeated measures was used to compare within-group responses to stimulus presentation and removal, and between-group differences were evaluated by testing for significant interactions between stimulus response and genotype.

Behaviour. All analysis was performed in ImageJ. Well-fed adult animals were briefly washed twice in NGM buffer and transferred to a 90-mm NGM agar Petri plate with no food. The entire plate was imaged for several minutes with a 5-megapixel camera (JAI) with a 60-mm lens (AF Micro-Nikkor; Nikon) using LabVIEW (National Instruments) at 2 frames per second. For each animal, 100 frames of video (50 s) were analysed. An ellipse was fitted to a spline corresponding to the animal's midline, and the aspect ratio of this ellipse was used as a size-independent index of the animal's posture, in which lower values indicate higher-amplitude and lower-wavelength gait. Frames in which animals were not performing forward locomotion (less than 2%) were excluded.

Connectomics. Network graphs were drawn with Cytoscape (<http://www.cytoscape.org>) on the basis of serial electron microscopy^{5,6}.

Butterfly genome reveals promiscuous exchange of mimicry adaptations among species

The *Heliconius* Genome Consortium*

The evolutionary importance of hybridization and introgression has long been debated¹. Hybrids are usually rare and unfit, but even infrequent hybridization can aid adaptation by transferring beneficial traits between species. Here we use genomic tools to investigate introgression in *Heliconius*, a rapidly radiating genus of neotropical butterflies widely used in studies of ecology, behaviour, mimicry and speciation^{2–5}. We sequenced the genome of *Heliconius melpomene* and compared it with other taxa to investigate chromosomal evolution in Lepidoptera and gene flow among multiple *Heliconius* species and races. Among 12,669 predicted genes, biologically important expansions of families of chemosensory and *Hox* genes are particularly noteworthy. Chromosomal organization has remained broadly conserved since the Cretaceous period, when butterflies split from the *Bombyx* (silkworm) lineage. Using genomic resequencing, we show hybrid exchange of genes between three co-mimics, *Heliconius melpomene*, *Heliconius timareta* and *Heliconius elevatus*, especially at two genomic regions that control mimicry pattern. We infer that closely related *Heliconius* species exchange protective colour-pattern genes promiscuously, implying that hybridization has an important role in adaptive radiation.

The butterfly genus *Heliconius* (Nymphalidae: Heliconiinae) is associated with a suite of derived life-history and ecological traits, including pollen feeding, extended lifespan, augmented ultraviolet colour vision, 'trap-lining' foraging behaviour, gregarious roosting and complex mating behaviours, and provides outstanding opportunities for genomic studies of adaptive radiation and speciation^{4,6}. The genus is best known for the hundreds of races with different colour patterns seen among its 43 species, with repeated examples of both convergent evolution among distantly related species and divergent evolution between closely related taxa³. Geographic mosaics of multiple colour-pattern races, such as in *Heliconius melpomene* (Fig. 1), converge to similar mosaics in other species, and this led to the hypothesis of mimicry². *Heliconius* are unpalatable to vertebrate predators and Müllerian mimicry of warning colour patterns enables species to share the cost of educating predators³. As a result of its dual role in mimicry and mate selection, divergence in wing pattern is also associated with speciation and adaptive radiation^{3,5}. A particularly recent radiation is the *melpomene*–silvaniform clade, in which mimetic patterns often seem to be polyphyletic (Fig. 1a). Most species in this clade occasionally hybridize in the wild with other clade members⁷. Gene genealogies at

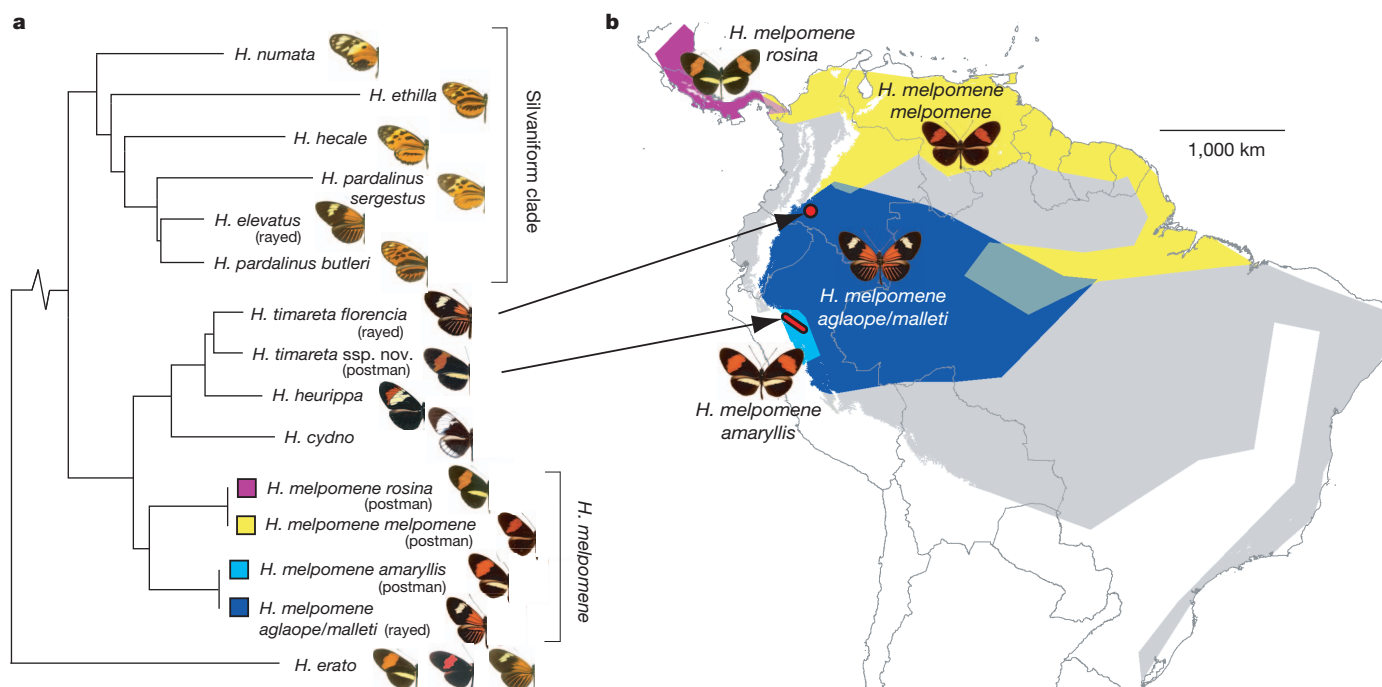


Figure 1 | Distribution, mimicry and phylogenetic relationships of sequenced taxa. **a**, Phylogenetic relationship of sequenced species and subspecies in the *melpomene*–silvaniform clade of *Heliconius*. *Heliconius elevatus* falls in the silvaniform clade, but it mimics colour patterns of *melpomene*–*timareta* clade taxa. Most other silvaniforms mimic unrelated ithomiine butterflies²⁴. **b**, Geographic distribution of postman and rayed

H. melpomene races studied here (blue, yellow and purple), and the entire distribution of *H. melpomene* (grey). The *H. timareta* races investigated have limited distributions (red) indicated by arrows and mimic sympatric races of *H. melpomene*. *Heliconius elevatus* and the other silvaniform species are distributed widely across the Amazon basin (Supplementary Information, section 22).

*Lists of participants and their affiliations appear at the end of the paper.

a small number of loci indicate introgression between species⁸, and one non-mimetic species, *Heliconius heurippa*, has a hybrid origin⁹. Adaptive introgression of mimicry loci is therefore a plausible explanation for parallel evolution of multiple mimetic patterns in the *melpomene*–*silvaniform* clade.

A *Heliconius melpomene melpomene* stock from Darién, Panama (Fig. 1), was inbred through five generations of sib mating. We sequenced a single male to $\times 38$ coverage (after quality filtering) using combined 454 and Illumina technologies (Supplementary Information, sections 1–8). The complete draft genome assembly, which is 269 megabases (Mb) in size, consists of 3,807 scaffolds with an N50 of 277 kb and contains 12,669 predicted protein-coding genes. Restriction-site-associated DNA (RAD) linkage mapping was used to assign and order 83% of the sequenced genome onto the 21 chromosomes (Supplementary Information, section 4). These data permit a considerably improved genome-wide chromosomal synteny comparison with the silkworm *Bombyx mori*^{10,11}.

Using 6,010 orthologues identified between *H. melpomene* and *B. mori*, we found that 11 of 21 *H. melpomene* linkage groups show homology to single *B. mori* chromosomes and that ten linkage groups have major contributions from two *B. mori* chromosomes (Fig. 2a and Supplementary Information, section 8), revealing several previously unidentified chromosomal fusions. These fusions on the *Heliconius* lineage most probably occurred after divergence from the sister genus *Eueides*⁴, which has the lepidopteran modal karyotype of $n = 31$ (ref. 12). Three chromosomal fusions are evident in *Bombyx* (*B. mori* chromosomes 11, 23 and 24; Fig. 2a), as required for evolution of the *Bombyx* $n = 28$ karyotype from the ancestral $n = 31$ karyotype. *Heliconius* and *Bombyx* lineages diverged in the Cretaceous, more than 100 million years ago¹¹, so the gross chromosomal structures of Lepidoptera genomes have remained highly conserved compared with those of flies or vertebrates^{13,14}. By contrast, small-scale rearrangements were frequent. In the comparison with *Bombyx*, we estimate there to be 0.05–0.13 breaks per megabase per million years, and in that with *Danaus plexippus* (Monarch butterfly), we estimate there to be 0.04–0.29 breaks per megabase per million years. Although lower than previously suggested for Lepidoptera¹⁵, these rates are comparable to those in *Drosophila* (Supplementary Information, section 8).

The origin of butterflies was associated with a switch from nocturnal to diurnal behaviour, and a corresponding increase in visual communication¹⁶. *Heliconius* have increased visual complexity through expression of a duplicate ultraviolet opsin⁶, in addition to the long-wavelength-, blue- and ultraviolet-sensitive opsins in *Bombyx*. We might therefore predict reduced complexity of olfactory genes, but in fact *Heliconius* and *Danaus*¹⁷ genomes have more chemosensory genes than any other insect genome: 33 and 34, respectively (Supplementary Information, section 9). For comparison, there are 24 in *Bombyx* and 3–4 in *Drosophila*¹⁸. Lineage-specific expansions of chemosensory genes were evident in both *Danaus* and *Heliconius* (Fig. 2b). By contrast, all three lepidopteran genomes have similar numbers of odorant binding proteins and olfactory receptors (Supplementary Information, section 9). *Hox* genes are involved in body plan development and show strong conservation across animals. We identified four additional *Hox* genes located between the canonical *Hox* genes *pb* and *zen*, orthologous to *shx* genes in *B. mori*¹⁹ (Supplementary Information, section 10). These *Hox* gene duplications in the butterflies and *Bombyx* have a common origin and are independent of the two tandem duplications known in dipterans (*zen2* and *bcd*). Immunity-related gene families are similar across all three lepidopterans (Supplementary Information, section 11), whereas there are extensive duplications and losses within dipterans²⁰.

The *Heliconius* reference genome allowed us to perform rigorous tests for introgression among *melpomene*–*silvaniform* clade species. We used RAD resequencing to reconstruct a robust phylogenetic tree based on 84 individuals of *H. melpomene* and its relatives, sampling on average 12 Mb, or 4%, of the genome (Fig. 1a and Supplementary Information, sections 12–18). We then tested for introgression between the sympatric co-mimetic postman butterfly races of *Heliconius melpomene amaryllis* and *H. timareta* ssp. nov. (Fig. 1) in Peru, using ‘ABBA/BABA’ single nucleotide sites and Patterson’s *D*-statistics (Fig. 3a), originally developed to test for admixture between Neanderthals and modern humans^{21,22} (Supplementary Information, section 12). Genome-wide, we found an excess of ABBA sites, giving a significantly positive Patterson’s *D* of 0.037 ± 0.003 (two-tailed *Z*-test for $D = 0$, $P = 1 \times 10^{-40}$), indicating greater genome-wide introgression between the sympatric mimetic taxa *H. melpomene amaryllis* and

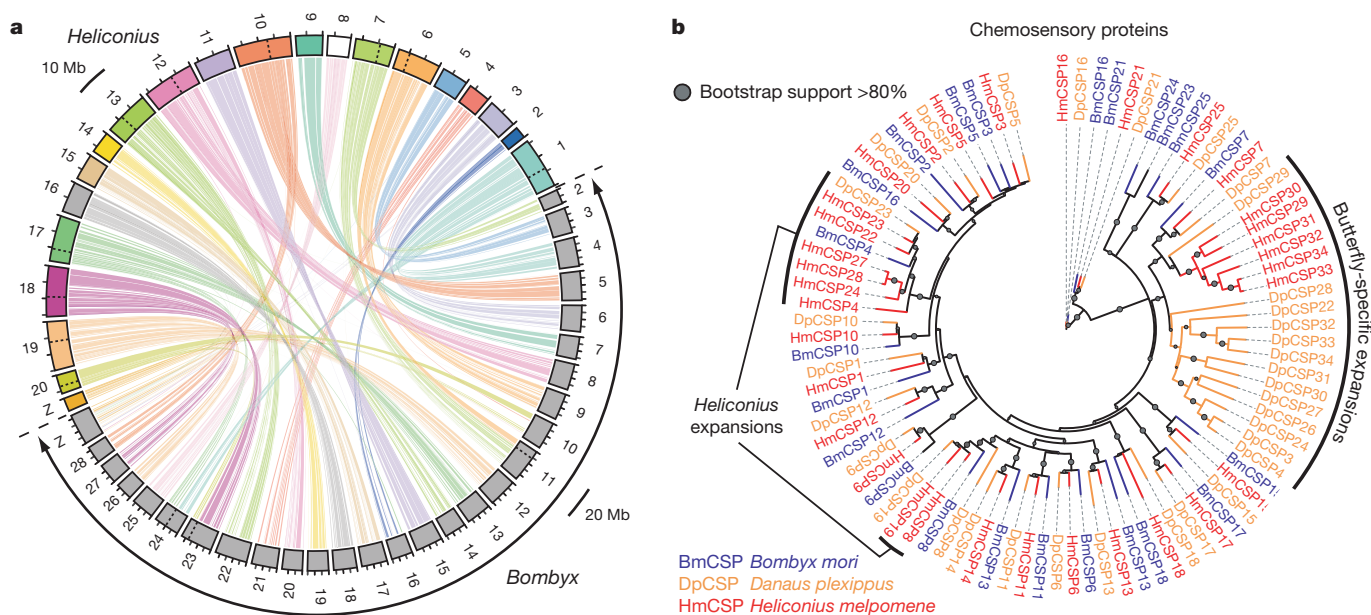


Figure 2 | Comparative analysis of synteny and expansion of the chemosensory genes. **a**, Maps of the 21 *Heliconius* chromosomes (colour) and of the 28 *Bombyx* chromosomes (grey) based on positions of 6,010 orthologue pairs demonstrate highly conserved synteny and a shared $n = 31$ ancestor

(Supplementary Information, section 8). Dotted lines within chromosomes indicate major chromosomal fusions. **b**, Maximum-likelihood tree showing expansions of chemosensory protein (CSP) genes in the two butterfly genomes.

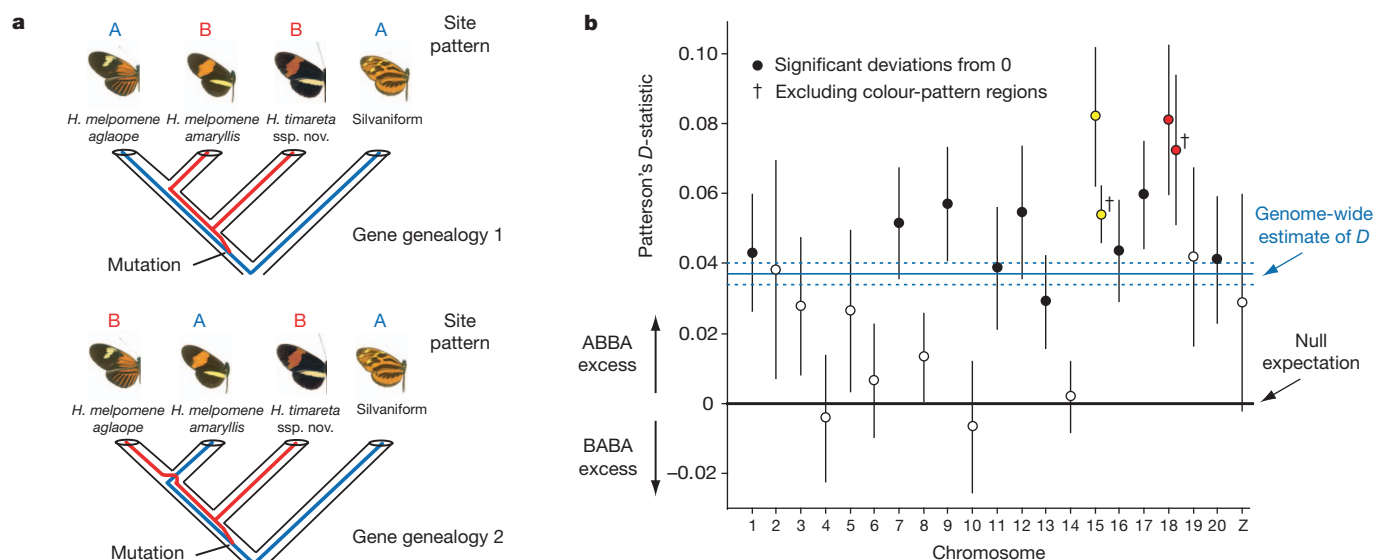


Figure 3 | Four-taxon ABBA/BABA test of introgression. **a**, ABBA and BABA nucleotide sites employed in the test are derived (—B—) in *H. timareta* compared with the silvaniform outgroup (—A—), but differ among *H. melpomene amaryllis* and *H. melpomene aglaope* (either ABBA or BABA). As this almost exclusively restricts attention to sites polymorphic in the ancestor of *H. timareta* and *H. melpomene*, equal numbers of ABBA and BABA sites are expected under a null hypothesis of no introgression²², as depicted in the two gene genealogies. **b**, Distribution among chromosomes of Patterson's

D-statistic (\pm s.e.), which measures excess of ABBA sites over BABA sites²², here for the comparison: *H. m. aglaope*, *H. m. amaryllis*, *H. timareta* ssp. nov., silvaniform. Chromosomes containing the two colour-pattern regions (*B/D*, red; *N/Yb*, yellow) have the two highest *D*-statistics; the combinatorial probability of this occurring by chance is 0.005. The excess of ABBA sites ($0 < D < 1$) indicates introgression between sympatric *H. timareta* and *H. m. amaryllis*.

H. timareta ssp. nov. than between *H. melpomene aglaope* and *H. timareta* ssp. nov., which do not overlap spatially (Fig. 1b). On the basis of these *D*-statistics, we estimate that 2–5% of the genome was exchanged²¹ between *H. timareta* and *H. melpomene amaryllis*, to the exclusion of *H. melpomene aglaope*. (Supplementary Information, section 12). Exchange was not random. Of the 21 chromosomes, 11 have significantly positive *D*-statistics, and the strongest signals of introgression were found on the two chromosomes containing known mimicry loci *B/D* and *N/Yb* (Fig. 3b and Supplementary Information, section 15).

Perhaps the best-known case of Müllerian mimicry is the geographic mosaic of ~30 bold postman and rayed colour-pattern races of *H. melpomene* (Fig. 1b and Supplementary Information, section 22), which mimic a near-identical colour-pattern mosaic in *Heliconius erato* (Fig. 1a), among other *Heliconius* species. Mimicry variation is mostly controlled by a few loci with strong effects. Mimetic pattern differences between the postman *H. m. amaryllis* and the rayed *H. m. aglaope* races studied here (Fig. 1a) are controlled by the *B/D* (red pattern) and *N/Yb* (yellow pattern) loci^{23,24}. These loci are located on the two chromosomes that show the highest *D*-statistics in our RAD analysis (Fig. 3b). To test whether mimicry loci might be introgressed between co-mimetic *H. timareta* and *H. melpomene*⁷ (Fig. 1a), we resequenced the colour-pattern regions *B/D* (0.7 Mb) and *N/Yb* (1.2 Mb), and 1.8 Mb of unlinked regions across the genome, from both postman and ray-patterned *H. melpomene* and *H. timareta* from Peru and Colombia, and six silvaniform outgroup taxa (Fig. 1a and Supplementary Information, section 12). To test for introgression at the *B/D* mimicry locus, we compared rayed *H. m. aglaope* and postman *H. m. amaryllis* as the ingroup with postman *H. timareta* ssp. nov. (Fig. 3a) and found large, significant peaks of shared, fixed ABBA nucleotide sites combined with an almost complete lack of BABA sites (Fig. 4b). This provides evidence that blocks of shared sequence variation in the *B/D* region were exchanged between postman *H. timareta* and postman *H. melpomene* in the genomic region known to determine red mimicry patterns between races of *H. melpomene*^{23,24} (Fig. 4a).

For a reciprocal test, we used the same *H. melpomene* races as the ingroup to compare with rayed *Heliconius timareta florenciae* at the

B/D region. In this case, correspondingly large and significant peaks of BABA nucleotide sites are accompanied by an almost complete absence of ABBA sites (Fig. 4c), indicating that variation at the same mimicry locus was also shared between rayed *H. timareta* and rayed *H. melpomene*. Equivalent results in the *N/Yb* colour-pattern region, controlling yellow colour-pattern differences, are in the expected directions for introgression and are highly significant for the test using postman *H. timareta* ssp. nov. ($P = 6 \times 10^{-34}$), but are not significant in rayed *H. t. florenciae* ($P = 0.13$; Supplementary Information, section 17). By contrast, hardly any ABBA or BABA sites are present in either comparison across 1.8 Mb in 55 genomic scaffolds that are unlinked to the colour-pattern regions (Supplementary Information, section 21). These concordant but reciprocal patterns of fixed ABBA and BABA substitutions occur almost exclusively within large genomic blocks at two different colour-pattern loci (449 and 99 sites for *B/D* and *N/Yb*, respectively; Fig. 4b, c and Supplementary Information, section 17). These patterns would be very hard to explain in terms of convergent functional-site evolution or random coalescent fluctuations. Instead, our results imply that derived colour-pattern elements have introgressed recently between both rayed and postman forms of *H. timareta* and *H. melpomene*.

To test whether colour-pattern loci might be shared more broadly across the clade, we used sliding-window phylogenetic analyses along the colour-pattern regions. For regions flanking and unlinked to colour-pattern loci, tree topologies are similar to the predominant signal recovered from the genome as a whole (Supplementary Information, section 18). Races of *H. melpomene* and *H. timareta* each form separate monophyletic sister groups and both are separated from the more distantly related silvaniform species (Fig. 4d). By contrast, topologies within the region of peak ABBA/BABA differences group individuals by colour pattern, and the species themselves become polyphyletic (Fig. 4e, f and Supplementary Information, sections 19 and 20). Remarkably, the rayed *H. elevatus*, a member of the silvaniform clade according to genome average relationships (Fig. 1a and Supplementary Information, section 18), groups with rayed races of unrelated *H. melpomene* and *H. timareta* in small sections within both *B/D* and *N/Yb* colour-pattern loci (Fig. 4e and Supplementary

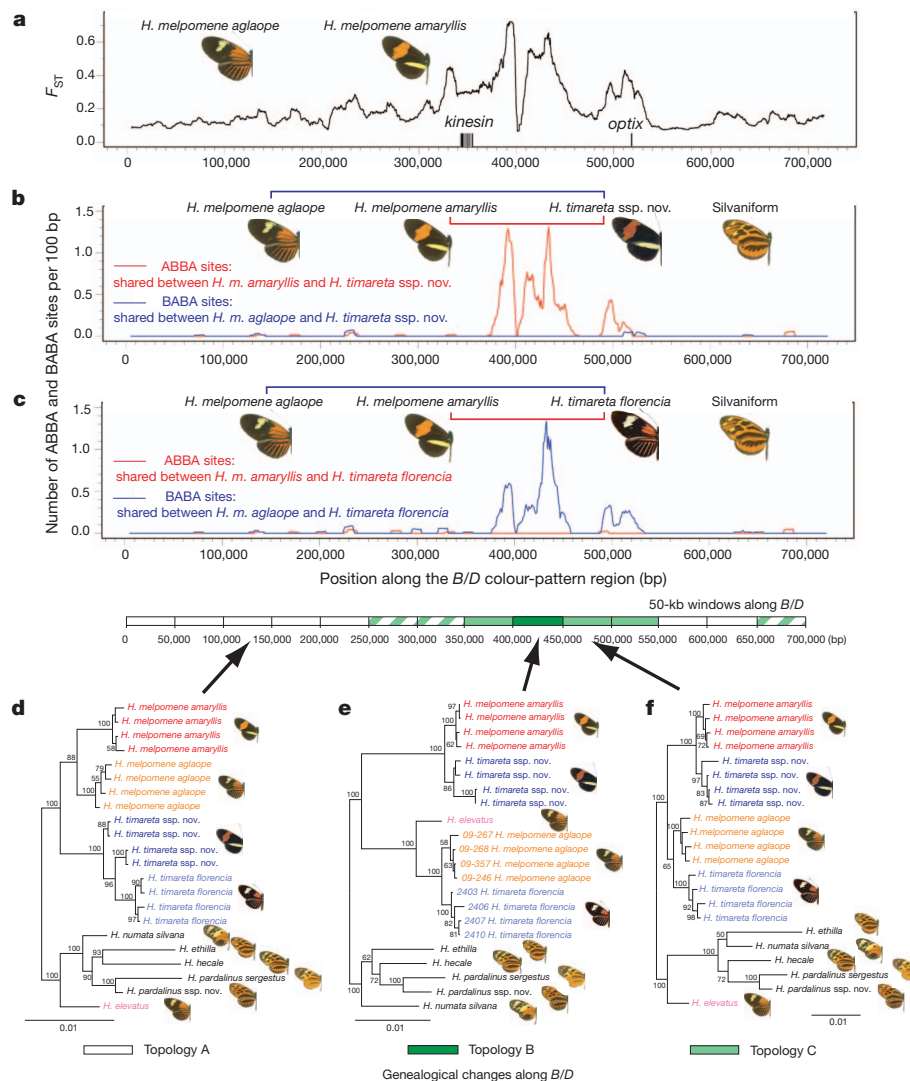


Figure 4 | Evidence for adaptive introgression at the B/D mimicry locus.

a, Genetic divergence between *H. melpomene* races *aglaope* (rayed) and *amaryllis* (postman) across a hybrid zone in northeast Peru. Divergence, F_{ST} , is measured along the B/D region (Supplementary Information 14) and peaks in the region known to control red wing pattern elements between the genes *kinesin* and *optix*²³. **b**, **c**, Distribution of fixed ABBA and BABA sites (see Fig. 3a) along B/D for two comparisons. Excesses of ABBA in **b** and BABA in **c** are highly significant (two-tailed Z-tests for $D = 0$; $D = 0.90 \pm 0.13$, $P = 5 \times 10^{-14}$ and $D = -0.91 \pm 0.10$, $P = 9 \times 10^{-24}$, respectively), indicating

Information, sections 19 and 20). These results are again most readily explained by introgression and fixation of mimicry genes.

We have developed a *de novo* reference genome sequence that will facilitate evolutionary and ecological studies in this key group of butterflies. We have demonstrated repeated exchange of large (~100-kb) adaptive regions among multiple species in a recent radiation. Our genome-scale analysis provides considerably greater power than previous tests of introgression^{8,25–27}. Our evidence suggests that *H. elevatus*, like *H. heurippa*⁹, was formed during a hybrid speciation event. The main genomic signal from this rayed species places it closest to *Heliconius pardalinus butleri* (Fig. 1a), but colour-pattern genomic regions resemble those of rayed races of *H. melpomene* (Fig. 4e and Supplementary Information, sections 18–21). Colour pattern is important in mating behaviour in *Heliconius*⁵, and the transfer of mimetic pattern may have enabled the divergent sibling species *H. elevatus* to coexist with *H. pardalinus* across the Amazon basin. Although it was long suspected that introgression might be important in evolutionary radiations¹, our results from the most diverse terrestrial

biome on the planet suggest that adaptive introgression is more pervasive than previously realized. The annotated genome version 1.1 is available on the *Heliconius* Genome Consortium's genome browser at <http://butterflygenome.org/> and this version will also be included in the next release of ENSEMBL Genomes. A full description of methods can be found in Supplementary Information.

Received 26 October 2011; accepted 12 March 2012.
Published online 16 May 2012.

Received 26 October 2011; accepted 12 March 2012.
Published online 16 May 2012.

1. Seehausen, O. Hybridization and adaptive radiation. *Trends Ecol. Evol.* **19**, 198–207 (2004).
2. Bates, H. W. Contributions to an insect fauna of the Amazon valley. Lepidoptera: Heliconiidae. *Trans. Linn. Soc. Lond.* **23**, 495–566 (1862).
3. Turner, J. R. G. Adaptation and evolution in *Heliconius*: a defense of neo-Darwinism. *Annu. Rev. Ecol. Syst.* **12**, 99–121 (1981).
4. Brown, K. S. The biology of *Heliconius* and related genera. *Annu. Rev. Entomol.* **26**, 427–457 (1981).
5. Jiggins, C. D., Naisbit, R. E., Coe, R. L. & Mallet, J. Reproductive isolation caused by colour pattern mimicry. *Nature* **411**, 302–305 (2001).

6. Briscoe, A. D. *et al.* Positive selection of a duplicated UV-sensitive visual pigment coincides with wing pigment evolution in *Heliconius* butterflies. *Proc. Natl Acad. Sci. USA* **107**, 3628–3633 (2010).
7. Mallet, J. in *Speciation and Patterns of Diversity* (eds Butlin, R. K., Schluter, D. & Bridle, J. R.) 177–194 (Cambridge Univ. Press, 2009).
8. Kronforst, M. R. Gene flow persists millions of years after speciation in *Heliconius* butterflies. *BMC Evol. Biol.* **8**, 98 (2008).
9. Salazar, C. *et al.* Genetic evidence for hybrid trait speciation in *Heliconius* butterflies. *PLoS Genet.* **6**, e1000930 (2010).
10. International Silkworm Genome Consortium. The genome of a lepidopteran model insect, the silkworm *Bombyx mori*. *Insect Biochem. Mol. Biol.* **38**, 1036–1045 (2008).
11. Pringle, E. G. *et al.* Synteny and chromosome evolution in the Lepidoptera: evidence from mapping in *Heliconius melpomene*. *Genetics* **177**, 417–426 (2007).
12. Robinson, R. *Lepidoptera Genetics* 557–598 (Pergamon, 1971).
13. Deng, Q., Zeng, Q., Qian, Y., Li, C. & Yang, Y. Research on the karyotype and evolution of the *Drosophila melanogaster* species group. *J. Genet. Genomics* **34**, 196–213 (2007).
14. Kemkemer, C. *et al.* Gene synteny comparisons between different vertebrates provide new insights into breakage and fusion events during mammalian karyotype evolution. *BMC Evol. Biol.* **9**, 84 (2009).
15. d'Alençon, E. *et al.* Extensive synteny conservation of holocentric chromosomes in Lepidoptera despite high rates of local genome rearrangements. *Proc. Natl Acad. Sci. USA* **107**, 7680–7685 (2010).
16. Vane-Wright, R. I. & Boppré, M. Visual and chemical signalling in butterflies: functional and phylogenetic perspectives. *Philos. Trans. R. Soc. Lond., B* **340**, 197–205 (1993).
17. Zhan, S., Merlin, C., Boore, J. L. & Reppert, S. M. The monarch butterfly genome yields insights into long-distance migration. *Cell* **147**, 1171–1185 (2011).
18. Vieira, F. G. & Rozas, J. Comparative genomics of the odorant-binding and chemosensory protein gene families across the Arthropoda: origin and evolutionary history of the chemosensory system. *Genome Biol. Evol.* **3**, 476–490 (2011).
19. Chai, C. L. *et al.* A genomewide survey of homeobox genes and identification of novel structure of the *Hox* cluster in the silkworm, *Bombyx mori*. *Insect Biochem. Mol. Biol.* **38**, 1111–1120 (2008).
20. Sackton, T. B. *et al.* Dynamic evolution of the innate immune system in *Drosophila*. *Nature Genet.* **39**, 1461–1468 (2007).
21. Green, R. E. *et al.* A draft sequence of the Neandertal genome. *Science* **328**, 710–722 (2010).
22. Durand, E. Y., Patterson, N., Reich, D. & Slatkin, M. Testing for ancient admixture between closely related populations. *Mol. Biol. Evol.* **28**, 2239–2252 (2011).
23. Reed, R. D. *et al.* *optix* drives the repeated convergent evolution of butterfly wing pattern mimicry. *Science* **333**, 1137–1141 (2011).
24. Nadeau, N. J. *et al.* Evidence for genomic islands of divergence among hybridizing species and subspecies of *Heliconius* butterflies obtained by large-scale targeted sequencing. *Phil. Trans. R. Soc. B* **367**, 343–353 (2012).
25. Bull, V. *et al.* Polyphyly and gene flow between non-sibling *Heliconius* species. *BMC Biol.* **4**, 11 (2006).
26. Kim, M. *et al.* Regulatory genes control a key morphological and ecological trait transferred between species. *Science* **322**, 1116–1119 (2008).
27. Song, Y. *et al.* Adaptive introgression of anticoagulant rodent poison resistance by hybridization between Old World mice. *Curr. Biol.* **21**, 1296–1301 (2011).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the governments of Colombia, Peru and Panama for permission to collect the butterflies. Sequencing was funded by contributions from consortium members. We thank M. Abanto for assistance in raising the inbred line. Individual laboratories were funded by the Leverhulme Trust (C.D.J.), the John Fell Fund and Christ Church College, Oxford (L.C.F.), The Royal Society (M.J., C.D.J.), the NSF (W.O.M., M.R.K., R.D.R., S.M., A.D.B.), the NIH (M.R.K., S.L.S., J.A.Y.), the CNRS (M.J.), the ERC (M.J., P.W.H.H.), the Banco de la República and COLCIENCIAS (M.L.) and the BBSRC (J.M., C.D.J., M.L.B. and R.H.F.-C.).

Author Contributions Consortium leaders: C.D.J., W.O.M. *Heliconius* Genome Consortium Principal Investigators: R.H.F.-C., M.R.K., M.J., J.M., S.M., R.D.R., M.L.B., L.E.G., M.L., G.L. Introgression study leader: J.M. Lead investigators: K.K.D., J.R.W., N.J.N., A.W., J.W.D., A.D.B., L.C.F., D.S.T.H., S.M., C.S., J.J.L., A.V.Z. Sequencing: S.R., S.E.S., A.L.B., M.T., K. Gharbi, C.E., M.L.B., R.A.G., Y.H., J.C.J., C.K., T.M., D.M.M., F.O., L.-L.P., J.Q., R.L.T., K.C.W., Y.-Q.W. Assembly: A.V.Z., J.A.Y., S.L.S., A.P., K. Gordon. RAD map and assembly verification: J.W.D., S.W.B., M.L.B., L.S.M., D.D.K., J.R.W., P.A.W. Geographic distribution map: N.R. Annotation: J.R.W., D.S.T.H., D.W., D.L., K.J.H., S.A., P.A.W., P.K. Genome browser and databases: D.S.T.H., J.J.L. Manual annotation and evolutionary analyses: A.D.B., E.J.-J., F.Y. (olfactory proteins); L.C.F., P.W.H.H., J.R.W. (*Hox* genes); A.S., T.D., D.M., S.M. (microRNAs); W.J.P., F.M.J. (immune genes); R.T.J., R.C. (P450 genes); H.V., S.-J.A., D.G.H. (uridine diphosphate glucuronosyltransferase genes); Y.P. (ribosomal proteins); S.W.B., M.L.B., A.D.B., N.L.C., B.A.C., L.C.F., H.M.H., C.D.J., F.M.J., M.J., D.D.K., M.R.K., J.M., A.M., S.P.M., N.J.N., W.J.P., R.P., M.A.S., A.T.-T., A.W., F.Y. (manual annotation group); B.A.C., D.A.R. (transposable elements); D.A.B. (orthologue predictions); A.W., J.W.D., D.G.H., K. Gordon (synteny); K.K.D., N.J.N., J.W.D., S.H.M., C.S., C.D.J., M.J., J.M. (introgression analysis). K.K.D. and J.R.W. contributed equally to this work.

Author Information The genome sequence has been submitted to the European Nucleotide Archive under accession numbers HE667773–HE672081. Additional short

read sequences have been submitted to the European Nucleotide Archive under accession numbers ERP000993 and ERP000991. Reprints and permissions information is available at www.nature.com/reprints. This paper is distributed under the terms of the Creative Commons Attribution-Non-Commercial-Share Alike licence, and is freely available to all readers at www.nature.com/nature. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to J.M. (jmallet@oeb.harvard.edu).

The *Heliconius* Genome Consortium Kanchon K. Dasmahapatra¹, James R. Walters², Adriana D. Briscoe³, John W. Davey⁴, Annabel Whibley⁵, Nicola J. Nadeau², Aleksey V. Zimin⁶, Daniel S. T. Hughes⁷, Laura C. Ferguson⁸, Simon H. Martin², Camilo Salazar^{2,9}, James J. Lewis³, Sebastian Adler¹⁰, Seung-Joon Ahn¹¹, Dean A. Baker¹², Simon W. Baxter¹³, Nicola L. Chamberlain¹³, Ritika Chauhan¹⁴, Brian A. Counterman¹⁵, Tamas Dalmay¹⁶, Lawrence E. Gilbert¹⁷, Karl Gordon¹⁸, David G. Heckel¹¹, Heather M. Hines¹⁹, Katharina J. Hoff¹⁰, Peter W. H. Holland⁸, Emmanuelle Jacquin-Joly²⁰, Francis M. Jiggins²¹, Robert T. Jones⁵, Durrell D. Kapan^{22,23}, Paul Kersey⁷, Gerardo Lamas²⁴, Daniel Lawson⁷, Daniel Mapleson²⁵, Luana S. Maroja²⁶, Arnaud Martin³, Simon Moxon²⁷, William J. Palmer²¹, Riccardo Papa²⁸, Alexie Papanicolaou¹⁸, Yannick Pauchet¹¹, David A. Ray^{29,30}, Neil Rosser¹, Steven L. Salzberg³¹, Megan A. Supple³², Alison Surridge², Ayse Tenger-Trolander¹³, Heiko Vogel¹¹, Paul A. Wilkinson³³, Derek Wilson⁷, James A. Yorke⁵, Furong Yuan⁵, Alexi L. Balmuth³⁴, Cathlene Eland³⁴, Karim Gharbi³⁴, Marian Thomson³⁴, Richard A. Gibbs³⁵, Yi Han³⁵, Joy C. Jayaseelan³⁵, Christie Kovar³⁵, Tittu Mathew³⁵, Donna M. Muzny³⁵, Fiona Onger³⁵, Ling-Ling Pu³⁵, Jiaxin Qu³⁵, Rebecca L. Thornton³⁵, Kim C. Worley³⁵, Yuan-Qing Wu³⁵, Mauricio Linares³⁶, Mark L. Blaxter^{4,34}, Richard H. ffrench-Constant¹⁴, Mathieu Joron⁵, Marcus R. Kronforst¹³, Sean P. Mullen³⁷, Robert D. Reed³, Steven E. Scherer³⁵, Stephen Richards³⁵, James Mallet^{1,38}, W. Owen McMillan⁹, Chris D. Jiggins^{2,9}

Affiliations for participants: ¹Department of Genetics, Evolution and Environment, University College London, Gower Street, London WC1E 6BT, UK. ²Department of Zoology, Downing Street, University of Cambridge, Cambridge CB2 3EJ, UK. ³Department of Ecology and Evolutionary Biology, University of California, Irvine, California 92697, USA. ⁴Institute of Evolutionary Biology, Ashworth Laboratories, University of Edinburgh, West Mains Road, Edinburgh EH9 3JT, UK. ⁵CNRS UMR 7205, Muséum National d'Histoire Naturelle, 45 rue Buffon, Paris 75005, France. ⁶Institute for Physical Science and Technology, University of Maryland, College Park, Maryland 20742, USA. ⁷European Bioinformatics Institute, Hinxton CB10 1SD, UK. ⁸Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK. ⁹Smithsonian Tropical Research Institute, Smithsonian Tropical Research Institute, Apartado Postal 0843-03092, Panamá, República de Panamá. ¹⁰Institut für Mathematik und Informatik, Universität Greifswald, 17487 Greifswald, Germany. ¹¹Max Planck Institute for Chemical Ecology, D-07745 Jena, Germany. ¹²Ecology and Evolution, Imperial College London, London SW7 2AZ, UK. ¹³FAS Center for Systems Biology, Harvard University, Cambridge, Massachusetts 02138, USA. ¹⁴Centre for Ecology and Conservation, School of Biosciences, University of Exeter, Penryn TR10 9EZ, UK. ¹⁵Department of Biology, Mississippi State University, Mississippi State, Mississippi 39762, USA. ¹⁶School of Biological Sciences, University of East Anglia, Norwich Research Park, Norwich NR4 7TJ, UK. ¹⁷Section of Integrative Biology and Brackenridge Field Laboratory, University of Texas, Austin, Texas 78712, USA. ¹⁸Black Mountain Laboratories, CSIRO Ecosystem Sciences, Clunies Ross Street, Canberra, Australian Capital Territory 2601, Australia. ¹⁹Department of Genetics, North Carolina State University, Raleigh, North Carolina 27695, USA. ²⁰UMR-A 1272 INRA-Université Pierre et Marie Curie, Physiologie de l'Insecte: Signalisation et Communication, Route de Saint-Cyr, Versailles Cedex 78026, France. ²¹Department of Genetics, Downing Street, University of Cambridge, Cambridge CB2 3EJ, UK. ²²Department of Entomology, Center for Comparative Genomics, California Academy of Sciences, 55 Music Concourse Drive, San Francisco, California 94118, USA. ²³Center for Conservation and Research Training, Pacific Biosciences Research Center, University of Hawaii at Manoa, 3050 Maile Way, Gilmore 406, Honolulu, Hawaii 96822, USA. ²⁴Museo de Historia Natural, Universidad Nacional Mayor de San Marcos, Apartado 14-0434, Lima, Peru. ²⁵School of Computing Sciences, University of East Anglia, Norwich Research Park, Norwich NR4 7TJ, UK. ²⁶Department of Biology, Williams College, Williamstown, Massachusetts 01267, USA. ²⁷Department of Genetics, Yale University School of Medicine, 333 Cedar Street, New Haven, Connecticut 06520, USA. ²⁸Department of Biology, University of Puerto Rico, PO Box 23360, Río Piedras, 00931-3360 Puerto Rico. ²⁹Department of Biochemistry, Molecular Biology, Entomology and Plant Pathology, Mississippi State University, Mississippi State, Mississippi 39762, USA. ³⁰Institute for Genomics, Biocomputing and Biotechnology, Mississippi State University, Mississippi State, Mississippi 39759, USA. ³¹McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University, Baltimore, Maryland 21205, USA. ³²Biomathematics Program, North Carolina State University, Raleigh, North Carolina 27695, USA. ³³School of Biological Sciences, University of Bristol, Bristol BS8 1UG, UK. ³⁴The GenePool, Ashworth Laboratories, University of Edinburgh, West Mains Road, Edinburgh EH9 3JT, UK. ³⁵Human Genome Sequencing Center, Baylor College of Medicine, One Baylor Plaza, Houston, Texas 77030, USA. ³⁶Facultad de Ciencias Naturales y Matemáticas, Universidad del Rosario and Instituto de Genética, Universidad de los Andes, Bogotá, Colombia. ³⁷Department of Biology, Boston University, 5 Cummington Street, Boston, Massachusetts 02215, USA. ³⁸Department of Organismic and Evolutionary Biology, Harvard University, 16 Divinity Avenue, Cambridge, Massachusetts 02138, USA.

Reduced airway surface pH impairs bacterial killing in the porcine cystic fibrosis lung

Alejandro A. Pezzullo^{1*}, Xiao Xiao Tang^{1,2*}, Mark J. Hoegger¹, Mahmoud H. Abou Alaiwa¹, Shyam Ramachandran³, Thomas O. Moninger⁴, Phillip H. Karp^{1,2}, Christine L. Wohlford-Lenane³, Henk P. Haagsman⁵, Martin van Eijk⁵, Botond Bánfi¹, Alexander R. Horswill⁶, David A. Stoltz¹, Paul B. McCray Jr³, Michael J. Welsh^{1,2,7} & Joseph Zabner¹

Cystic fibrosis (CF) is a life-shortening disease caused by mutations in the cystic fibrosis transmembrane conductance regulator (CFTR) gene¹. Although bacterial lung infection and the resulting inflammation cause most of the morbidity and mortality, how the loss of CFTR function first disrupts airway host defence has remained uncertain^{2–6}. To investigate the abnormalities that impair elimination when a bacterium lands on the pristine surface of a newborn CF airway, we interrogated the viability of individual bacteria immobilized on solid grids and placed onto the airway surface. As a model, we studied CF pigs, which spontaneously develop hallmark features of CF lung disease^{7,8}. At birth, their lungs lack infection and inflammation, but have a reduced ability to eradicate bacteria⁸. Here we show that in newborn wild-type pigs, the thin layer of airway surface liquid (ASL) rapidly kills bacteria *in vivo*, when removed from the lung and in primary epithelial cultures. Lack of CFTR reduces bacterial killing. We found that the ASL pH was more acidic in CF pigs, and reducing pH inhibited the antimicrobial activity of ASL. Reducing ASL pH diminished bacterial killing in wild-type pigs, and, conversely, increasing ASL pH rescued killing in CF pigs. These results directly link the initial host defence defect to the loss of CFTR, an anion channel that facilitates HCO₃[−] transport^{9–13}. Without CFTR, airway epithelial HCO₃[−] secretion is defective, the ASL pH falls and inhibits antimicrobial function, and thereby impairs the killing of bacteria that enter the newborn lung. These findings suggest that increasing ASL pH might prevent the initial infection in patients with CF, and that assaying bacterial killing could report on the benefit of therapeutic interventions.

Proposed origins of CF lung disease include reduced mucociliary clearance due to defects such as decreased ASL volume or altered mucus, reduced bacterial killing by ASL antimicrobials, defective bacterial elimination by phagocytes, abnormal inflammatory responses, and reduced or increased bacterial binding by airway epithelia^{2–6}. One or more of these defects could be responsible. Two factors have made it difficult to distinguish among hypotheses and identify the initiating insults. First, as with many diseases, the clinical manifestations may not reflect the original defects, but it is problematic to study CF at its onset in newborn human infants. Second, mice with disrupted *Cftr* genes do not develop typical CF¹⁴.

To circumvent these obstacles, we generated *CFTR*^{−/−} pigs (CF pigs)⁷. Within months of birth, CF pigs spontaneously develop hallmark features of CF lungs such as airway inflammation, infection, tissue remodelling, mucus accumulation and airway obstruction^{7,8}. Although at birth they exhibit none of these features, they already manifest a host defence defect against bacteria. Thus, newborn CF pigs provide an unprecedented opportunity to investigate mechanisms

impairing host defence and initiating disease, because they allow CF:non-CF comparisons without secondary confounders.

In previous work, we instilled *Staphylococcus aureus* into airways and four hours later found more bacteria in CF than in non-CF pigs⁸. However, that study revealed little about the responsible mechanisms; we do not know whether bacteria were removed or killed within the lung, whether bacteria grew after instillation, whether phagocytic cells eliminated bacteria, whether bacteria bound to surfaces, or whether deposition and sampling were identical in all animals.

To investigate initial host defence defects, we developed a simple assay that tested viability of individual bacteria. We chemically linked biotin to *S. aureus*, bound streptavidin to gold grids, and combined them to attach *S. aureus* to the grids (Fig. 1a, b). We chose *S. aureus* because we frequently isolate it from porcine CF lungs, and it is the most common organism isolated from young children with CF^{8,15}. A fluorescent live/dead stain showed the state of bacteria. Exposing grids to ethanol killed most *S. aureus* (Fig. 1c). Importantly, placing grids on the porcine tracheal surface *in vivo* also killed bacteria.

In 6–15-h-old pigs, we made a small tracheal incision and placed bacteria-coated grids on the airway surface. Even 30-s applications on non-CF airways killed *S. aureus* (Fig. 1d). Applying grids to littermate CF pigs killed approximately half as many bacteria. We administered methacholine to stimulate the secretion of submucosal glands, which produce substantial amounts of antimicrobials^{16,17}, and to allow us to collect ASL for other studies. After methacholine application, the CF:non-CF differences persisted (Fig. 1e). We predicted that antimicrobial activity would also be detected if we removed methacholine-stimulated ASL and studied it with conventional colony-forming unit (CFU) assays. Indeed, bacterial killing was reduced in CF secretions (Fig. 1f).

We also applied *S. aureus*-coated grids to primary cultures of porcine airway epithelia and found reduced killing in CF (Fig. 1g). Previous data suggest that the host defence defect involves many different bacteria^{8,15}. Therefore, we tested *Pseudomonas aeruginosa*-coated grids and found defective killing by CF epithelia (Fig. 1h). We also added *S. aureus* directly to cultured epithelia. Most non-CF epithelia eliminated low inocula of bacteria, but bacteria grew on most CF epithelia (Fig. 1i). At the highest inocula, *S. aureus* infected both CF and non-CF epithelia.

These data indicate that ASL rapidly kills bacteria, and that CF impairs killing. The defect was partial, as CF ASL retained some activity. The assays allow several conclusions. Defective bacterial killing was not due to dysfunctional mucociliary clearance or abnormal killing by phagocytes; neither would explain the results with grids *in vivo* or studies of cultured epithelia. Also, we cannot attribute CF:non-CF differences to altered bacterial–epithelial binding, because we saw the difference with bacteria attached to grids and with ASL studied

¹Department of Medicine, Roy J. and Lucille A. Carver College of Medicine, University of Iowa, Iowa City, Iowa 52242, USA. ²Howard Hughes Medical Institute, Roy J. and Lucille A. Carver College of Medicine, University of Iowa, Iowa City, Iowa 52242, USA. ³Department of Pediatrics, Roy J. and Lucille A. Carver College of Medicine, University of Iowa, Iowa City, Iowa 52242, USA. ⁴The Central Electron Microscopy Facility, Roy J. and Lucille A. Carver College of Medicine, University of Iowa, Iowa City, Iowa 52242, USA. ⁵Department of Infectious Diseases and Immunology, Faculty of Veterinary Medicine, Utrecht University, Utrecht 3584 CL, The Netherlands. ⁶Department of Microbiology, Roy J. and Lucille A. Carver College of Medicine, University of Iowa, Iowa City, Iowa 52242, USA. ⁷Department of Molecular Physiology and Biophysics, Roy J. and Lucille A. Carver College of Medicine, University of Iowa, Iowa City, Iowa 52242, USA.

*These authors contributed equally to this work.

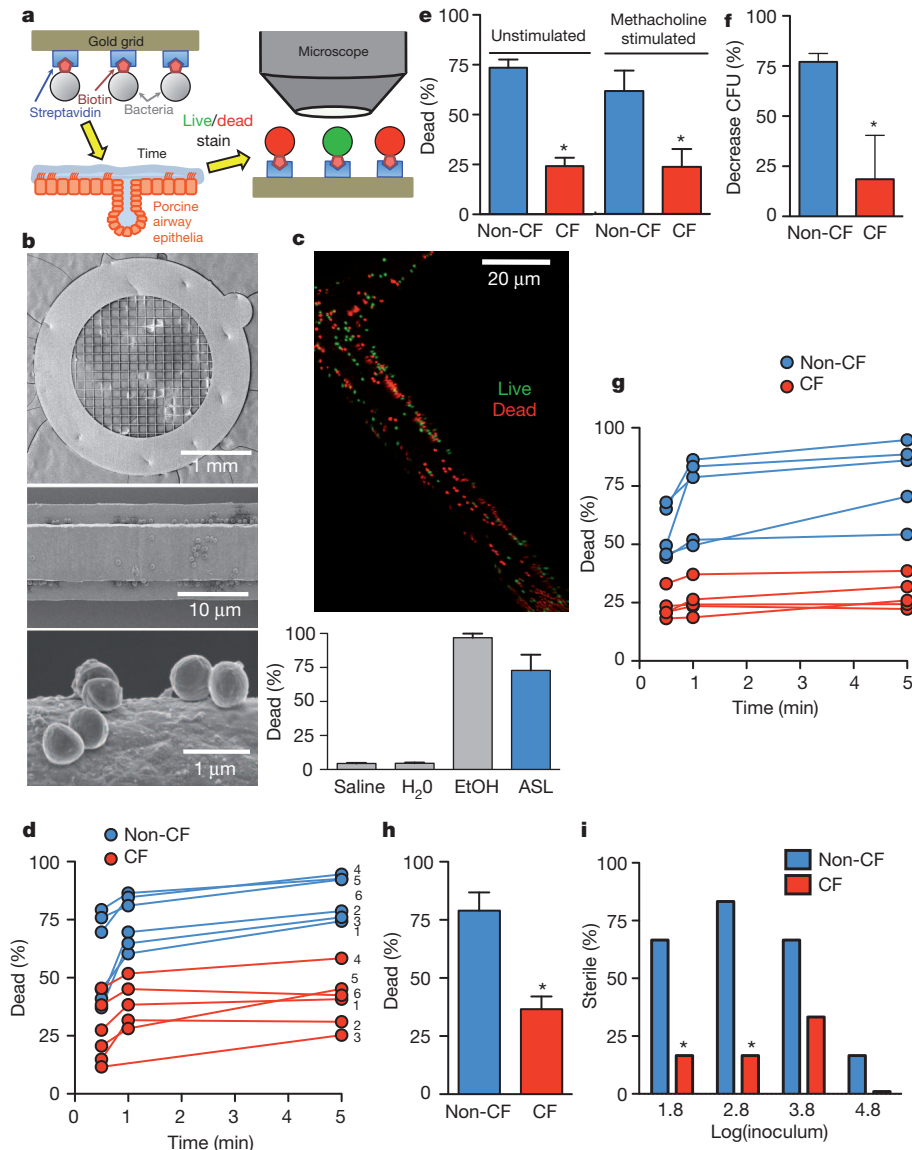


Figure 1 | Bacterial killing is impaired in CF ASL. **a**, Schematic showing biotin/streptavidin linking *S. aureus* to gold grids that were placed on the airway surface. After removal, bacteria were exposed to fluorescent live/dead stain (SYTO 9/propidium iodide), imaged and counted. **b**, Scanning electron photomicrographs of bacteria-coated grid (top), grid bar (middle) and individual bacteria (bottom). **c**, Bacteria-coated grid (green = live, red = dead) after placement for 5 min on the tracheal surface of 1-month-old, wild-type pig. Bottom shows percentage of bacteria that were dead after immersion in saline, water, 70% ethanol (EtOH) or placement on the tracheal surface. $n = 3$ each. **d**, *S. aureus*-coated grids were placed on the tracheal surface of newborn CF and non-CF pigs for indicated times. Data are percentage of dead bacteria. Each set of three time points is from a single animal. Numbers (1–6) indicate littermates; there is no 1-min time point for the CF pig in litter 3 owing to experimental error. For each pig, 2–3 grids were used at each time point, 5–16 fields were counted per grid, each field contained ~100–1,000 bacteria (Supplementary Fig. 6), and data from each field were averaged. Operators were blinded to genotype. CF was different from non-CF at all time points, $P < 0.01$.

ex vivo. Our earlier finding that newborn CF airways lack inflammation⁸ and the killing defect in cultured epithelia indicate that abnormal inflammation was not responsible. Furthermore, our bacteria-coated grid method also excludes differences in bacterial delivery, sampling or growth. Therefore, we reasoned that defective killing arose from either reduced amounts of ASL antimicrobial factors or inhibition of their function.

e, *S. aureus*-coated grids were placed for 30 s on the tracheal surface before and ~5 min after methacholine-stimulated secretion. $n = 6$ per genotype; $*P < 0.02$. For each genotype, data with and without methacholine do not differ significantly. **f**, ASL was removed from methacholine-stimulated pigs, bacteria (1×10^6 CFU ml^{-1}) were incubated in 1 μl ASL for 10 min in a micro-CFU assay and counted by dilution plating. $n = 6$ per genotype; $*P < 0.05$. **g**, *S. aureus*-coated grids were placed on the surface of primary epithelial cultures for indicated times, and the percentage of dead bacteria was determined. Each set of data points represents mean from epithelia from a different animal. CF was different from non-CF at all time points, $P < 0.001$. **h**, *P. aeruginosa*-coated grids were placed on the surface of cultured epithelia for 5 min. $n = 5$ cultures from different pigs per genotype. $*P < 0.001$. **i**, *S. aureus* in 100 nl H_2O were applied to the apical surface of cultured epithelia. The apical surface was washed 24 h later and bacteria counted by dilution plating. Data are percentage of epithelia that showed no bacterial growth. $n = 18$ per genotype; $*P < 0.005$, Fisher's exact test. All error bars denote s.e.m.

We investigated antimicrobials by measuring messenger RNA, protein and aggregate activity under optimal conditions. The abundance of transcripts for secreted antimicrobial proteins (Supplementary Table 1 and Supplementary Fig. 1) and proteins with known host defence functions showed no consistent differences between genotypes (Supplementary Table 2). In methacholine-stimulated ASL, concentrations of the two most abundant antimicrobials, lysozyme and

lactoferrin, as well as palate, lung and nasal epithelial clone (PLUNC) and surfactant protein SP-A, did not differ by genotype (Fig. 2a). To assay aggregate ASL antimicrobial function, we performed four experiments in which we maximized activity by reducing the ionic strength close to zero^{16,18,19}. First, we added isotonic, salt-free buffer to the apical surfaces of cultured airway epithelia. Under these control conditions, both genotypes showed equivalent killing of bacteria on grids (Fig. 2b). Second, ASL removed with water from cultured CF and non-CF epithelia killed bacteria to the same extent (Fig. 2c and Supplementary Fig. 2). Third, ASL removed from pigs and diluted 1:100 with water showed genotype-independent killing (Fig. 2d). Fourth, radial diffusion assays with 10 mM sodium phosphate in 1% agarose revealed areas of clearance for *S. aureus* and *Escherichia coli* that were similar for both genotypes (Fig. 2e). These data indicate that non-CF and CF ASL had similar amounts of antimicrobials. Thus, they suggested that CF:non-CF bacterial killing disparities derived from other differences in ASL composition.

An increased ionic strength inhibits activity of many antimicrobials^{16,18,19}. Studies of human CF airway epithelia in culture or xenografts reported either higher^{20,21} or the same^{22,23} ASL NaCl concentrations as non-CF epithelia, but an *in vivo* study²⁴ reported similar concentrations. Therefore, we measured Na⁺ and K⁺ concentrations in ASL collected from newborn pigs and found that they did not differ by genotype (Fig. 3a). In addition, ASL collected after methacholine stimulation showed similar ion concentrations to those measured under basal conditions, and only minor differences in K⁺ concentration between CF and non-CF ASL (Fig. 3b). Thus, different ASL Na⁺ and K⁺ concentrations do not explain the defective bacterial killing in CF.

Earlier studies indicated that pH can affect antimicrobial activity^{19,25}. Human and porcine airway epithelia exhibit CFTR-dependent HCO₃⁻ secretion^{11,12}, and CF reduces ASL pH^{26,27}. To assess ASL pH *in vivo*, we placed a planar pH-sensitive probe on the tracheal surface. pH was lower in CF than in non-CF ASL (Fig. 3c). Methacholine-stimulated ASL removed from CF pigs and measured with an optical pH probe was more acidic than ASL from non-CF pigs (Fig. 3d); pH

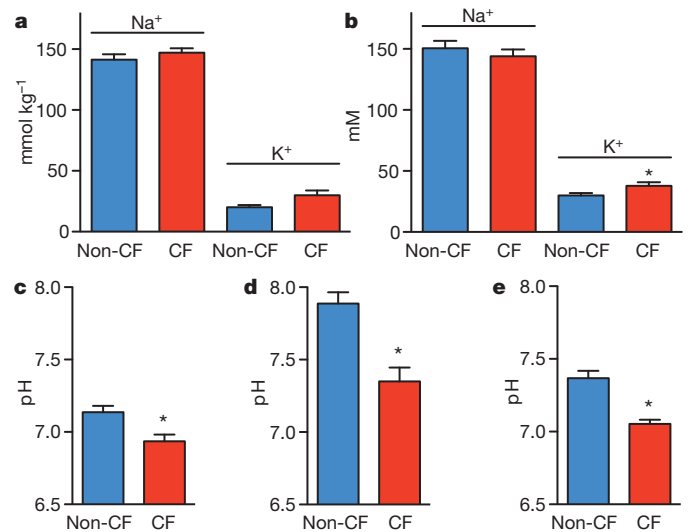


Figure 3 | ASL pH is more acidic in CF than non-CF. **a**, ASL was collected under basal conditions from the tracheal surface using Parafilm-coated paper; Na⁺ and K⁺ concentrations were measured as described in Methods (Supplementary Table 3). *n* = 8 non-CF and 6 CF pigs. **b**, ASL Na⁺ and K⁺ concentrations in methacholine-stimulated ASL. *n* = 16 non-CF and 14 CF pigs; **P* < 0.05. **c**, ASL pH was measured *in vivo* using a pH-sensitive planar optical probe placed on the tracheal surface. *n* = 6 non-CF and 7 CF pigs; littermates were used with one extra CF; **P* < 0.05. Studies were done with animals in 5% CO₂; therefore, the ASL CO₂ concentration was probably >5% owing to CO₂ production by the pigs. **d**, Methacholine-stimulated ASL was removed, and pH was measured with a micro-optical pH probe. pH was measured 10 min after removal in ambient CO₂, which probably increased pH values compared with *in vivo*. *n* = 10 pigs per genotype; littermate controls were used; **P* < 0.0005. **e**, ASL pH was measured in cultured airway epithelia using a fluorescent pH indicator. *n* = 5 epithelia per genotype, each from a different pig; **P* < 0.01. Calculated HCO₃⁻ concentrations using measured pH, the 5% CO₂ concentration, and Henderson-Hasselbalch equation were non-CF 28.1 ± 4.2 mM (*n* = 8) and CF 13.1 ± 2.4 mM (*n* = 8); *P* = 0.007. All error bars denote s.e.m.

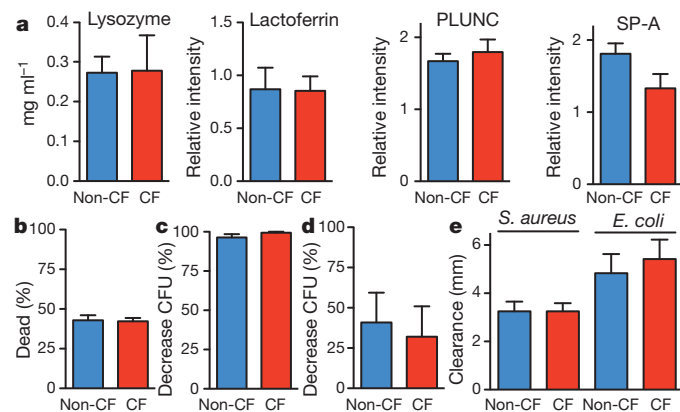


Figure 2 | CF and non-CF ASL have similar antimicrobial concentrations and aggregate antimicrobial activity under optimal conditions. **a**, Lysozyme concentration was measured with lysoplates; *n* = 8 per genotype. Quantitative western blots assayed lactoferrin, PLUNC and SP-A; data are relative intensity of blots, *n* = 12 per genotype. **b**, Isotonic xylitol (60 μl) was added to the apical surface of airway cultures (diluting ASL ~100:1). Three minutes later, two *S. aureus*-coated grids were placed on the epithelial surface for 1 min, then removed and counted. *n* = 6 epithelia per genotype, each from different pigs. **c**, ASL was removed from cultured epithelia by washing with 100 μl H₂O. ASL was incubated with *S. aureus* (3.3 × 10³ CFU ml⁻¹) for 60 min, and micro-CFU assays were used to measure antimicrobial activity. *n* = 12 per genotype. See also Supplementary Figure 2. **d**, Methacholine-stimulated ASL was diluted 1:100 in H₂O, incubated with *S. aureus* (1 × 10⁶ CFU ml⁻¹) for 60 min, and CFU assays were used to measure antimicrobial activity. *n* = 5 per genotype. **e**, Radial diffusion assays were used to measure antimicrobial activity of ASL collected from epithelial cultures. *n* = 6 per genotype. All error bars denote s.e.m.

was measured after removal in ambient CO₂, probably contributing to the higher absolute pH values. We also measured ASL pH in primary airway epithelial cultures using a fluorescent pH indicator and found reduced pH in CF (Fig. 3e). Although absolute pH values varied in different preparations, in all three, CF ASL was more acidic.

We tested whether pH affects ASL antimicrobial activity by removing ASL from newborn non-CF pigs, adjusting the pH and applying *S. aureus*-coated grids. Killing was pH-dependent, increasing as pH increased (Fig. 4a and Supplementary Fig. 3). We also tested lysozyme and lactoferrin; increasing pH increased *S. aureus* and *E. coli* killing (Fig. 4b and Supplementary Figs 4 and 5).

If pH is responsible for the differences in bacterial killing, we predicted that reducing ASL pH would inhibit bacterial killing in wild-type pigs, and raising pH would enhance killing in CF pigs. In non-CF pigs, increasing airway CO₂ reduced ASL pH and inhibited bacterial killing (Fig. 4c). In CF pigs, we aerosolized NaHCO₃ into the trachea. Compared with NaCl, NaHCO₃ increased ASL pH and enhanced killing (Fig. 4d).

Our results directly link CFTR mutations to defective bacterial eradication. CFTR is an anion channel that conducts HCO₃⁻ and works together with Cl⁻/HCO₃⁻ exchangers and H⁺ secretion to regulate ASL pH^{2,9,10,13}; its loss prevents airway epithelia from secreting HCO₃⁻ (refs 11, 12). The resulting decreased ASL pH inhibits antimicrobial function, thereby impairing killing of bacteria that enter the lung. Our findings with bacteria-coated grids *in vivo*, ASL removed from pigs, and primary epithelial cultures all point to this defect.

What about other defects that might commence CF lung disease? Progression from the pristine lung of a newborn to the chronically

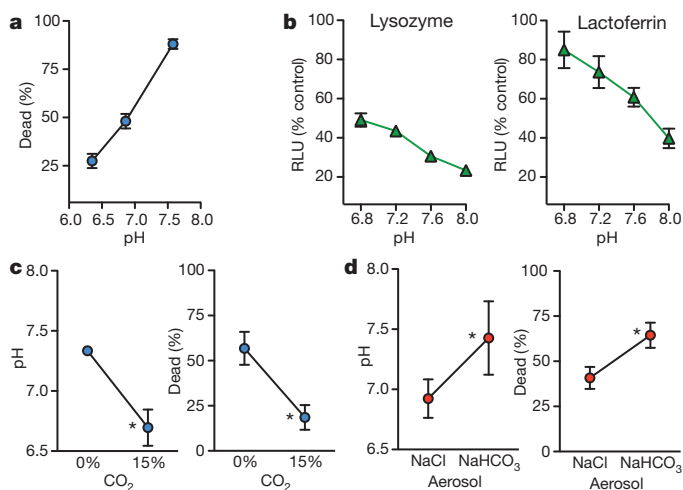


Figure 4 | Increasing ASL pH enhances antimicrobial activity.

a, Methacholine-stimulated ASL was removed from non-CF pigs, pH was adjusted with HCl, and antimicrobial activity was measured with *S. aureus*-coated grids. $n = 6$. See Supplementary Fig. 3 and Supplementary Table 3. **b**, Effect of pH on antimicrobial activity of 1.25 mg ml^{-1} lysozyme and 4 mg ml^{-1} lactoferrin in *S. aureus* luminescence assay. Data are relative luminescence compared with control at 30 min for lysozyme and 60 min for lactoferrin (Supplementary Fig. 5). Similar results were obtained with *E. coli* (Supplementary Fig. 4). RLU, relative light units. **c**, Tracheas of non-CF pigs were exposed to 0% or 15% CO_2 *in vivo*. pH was measured with a pH-sensitive planar electrode. $n = 6$ per genotype. Bacterial killing was measured with *S. aureus*-coated grids placed on the surface for 30 s. $n = 4$ per genotype; $*P < 0.05$. **d**, NaHCO_3 or NaCl (50 μl , 100 mM) was aerosolized onto the airway surface of CF pigs. pH ($n = 6$ per genotype) and bacterial killing ($n = 5$ per genotype) were measured as in **c**. $*P < 0.05$. All error bars denote s.e.m.

infected, inflamed, remodelled and obstructed lung of a person with advanced CF entails many steps. Our findings suggest that reduced antibacterial activity may initiate this downward spiral. However, we cannot exclude other abnormalities at the onset of disease. These might include abnormal mucociliary clearance, bacterial binding, inflammation or phagocytic function. As disease progresses, the relative importance of various defects may also change. As an example, reduced ASL pH or HCO_3^- might alter mucus secretion² and thereby impair mucociliary clearance either at the genesis of disease or as mucus secretion increases with inflammation and remodelling. Furthermore, ASL contains a complex mixture of peptides, proteins and lipids with individual^{16,28,29} and synergistic antimicrobial effects^{29,30}. Our data do not assess the relative importance or abundance of each factor or CF/non-CF differences.

Several potential fates await bacteria that land on the airway surface. They might be killed. They might remain metabolically active, but unable to reproduce. They might replicate. And/or they might be removed. The balance between these options and their timing determines whether or not infection ensues. Our grid assay tested one of these alternatives independently of the others, and the data indicate that ASL killed many, but not all bacteria. The speed of killing was notable, although killing might have continued after grids were withdrawn from ASL, rinsed and placed in indicator solution. Quick action is consistent with rapid bacterial permeabilization by antimicrobial proteins¹⁹. Our other assays showing reduced numbers of growing bacteria (CFUs) after being placed on epithelial cultures or in ASL reflect consequences of immediate killing plus potential effects of ASL on replication. Results from our earlier delivery of bacteria into porcine airways reflect all these host defences plus bacterial removal and phagocytosis⁸. Although several processes protect lungs when bacteria enter, ASL antimicrobials may be key to rapidly reducing numbers of viable organisms and thereby decreasing the probability that they will replicate, escape other mechanisms and colonize the lung.

Aerosolizing HCO_3^- onto CF airways *in vivo* increased bacterial killing. Our mechanistic findings and this result suggest that correcting ASL pH might prevent the initial CF infection. That might be accomplished by delivering HCO_3^- into airways, altering pH regulation by airway epithelia, enhancing activity of ASL antimicrobials, delivering pH-insensitive antimicrobials, or targeting mutant CFTR. These results also suggest that adapting the bacteria-coated grid method to assay bacterial killing in patients could be useful for assessing potential therapies. We also speculate that increasing ASL pH might prevent or treat airway infections in other diseases.

METHODS SUMMARY

Bacteria-coated grids were prepared by chemically modifying gold electron microscopy grids (200 mesh) and coating them with streptavidin. Bacteria (*S. aureus* isolate 43SA or *P. aeruginosa* PAO1) were cultured to log-phase growth and mixed with *N*-hydroxysulphosuccinimide-biotin. Streptavidin-coated grids were then incubated with biotinylated bacteria and rinsed in PBS. Bacteria-coated grids were placed on airway surfaces, removed, immediately rinsed with PBS, and then immersed in PBS containing the fluorescent indicators SYTO 9 and propidium iodide (Live/Dead Bacterial Viability Assay, Invitrogen) to assess viability. After 15 min, grids were rinsed with PBS and imaged with a laser-scanning confocal microscope (Olympus). The numbers of live and dead bacteria were counted in multiple fields by operators blinded to genotype. pH measurements were done *in vivo* using non-invasive dual lifetime referencing and a planar pH-sensitive optode. A needle-type fibre optic pH meter was used for pH measurements *ex vivo*. A ratiometric pH indicator was used in primary cultures of airway epithelia.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 3 January; accepted 5 April 2012.

- Welsh, M. J., Ramsey, B. W., Accurso, F. & Cutting, G. R. in *The Metabolic and Molecular Basis of Inherited Disease* (eds Scriver, C. R. et al.) 5121–5189 (McGraw-Hill, 2001).
- Quinton, P. M. Role of epithelial HCO_3^- transport in mucin secretion: lessons from cystic fibrosis. *Am. J. Physiol. Cell Physiol.* **299**, C1222–C1233 (2010).
- Gibson, R. L., Burns, J. L. & Ramsey, B. W. Pathophysiology and management of pulmonary infections in cystic fibrosis. *Am. J. Respir. Crit. Care Med.* **168**, 918–951 (2003).
- Wine, J. J. The genesis of cystic fibrosis lung disease. *J. Clin. Invest.* **103**, 309–312 (1999).
- Guggino, W. B. Cystic fibrosis and the salt controversy. *Cell* **96**, 607–610 (1999).
- Machen, T. E. Innate immune response in CF airway epithelia: hyperinflammatory? *Am. J. Physiol. Cell Physiol.* **291**, C218–C230 (2006).
- Rogers, C. S. et al. Disruption of the *CFTR* gene produces a model of cystic fibrosis in newborn pigs. *Science* **321**, 1837–1841 (2008).
- Stoltz, D. A. et al. Cystic fibrosis pigs develop lung disease and exhibit defective bacterial eradication at birth. *Sci. Transl. Med.* **2**, 29ra31 (2010).
- Fischer, H. & Widdicombe, J. H. Mechanisms of acid and base secretion by the airway epithelium. *J. Membr. Biol.* **211**, 139–150 (2006).
- Poulsen, J. H., Fischer, H., Illek, B. & Machen, T. E. Bicarbonate conductance and pH regulatory capability of cystic fibrosis transmembrane conductance regulator. *Proc. Natl Acad. Sci. USA* **91**, 5340–5344 (1994).
- Smith, J. J. & Welsh, M. J. cAMP stimulates bicarbonate secretion across normal, but not cystic fibrosis airway epithelia. *J. Clin. Invest.* **89**, 1148–1153 (1992).
- Chen, J.-H. et al. Loss of anion transport without increased sodium absorption characterizes newborn porcine cystic fibrosis airway epithelia. *Cell* **143**, 911–923 (2010).
- Garnett, J. P. et al. Novel role for pendrin in orchestrating bicarbonate secretion in cystic fibrosis transmembrane conductance regulator (CFTR)-expressing airway serous cells. *J. Biol. Chem.* **286**, 41069–41082 (2011).
- Grubb, B. R. & Boucher, R. C. Pathophysiology of gene-targeted mouse models for cystic fibrosis. *Physiol. Rev.* **79**, S193–S214 (1999).
- Foundation, C. F. Cystic Fibrosis Foundation Patient Registry: annual data report (2010).
- Ganz, T. Antimicrobial polypeptides. *J. Leukoc. Biol.* **75**, 34–38 (2004).
- Wine, J. J. & Joo, N. S. Submucosal glands and airway defense. *Proc. Am. Thorac. Soc.* **1**, 47–53 (2004).
- Smith, J. J., Travis, S. M., Greenberg, E. P. & Welsh, M. J. Cystic fibrosis airway epithelia fail to kill bacteria because of abnormal airway surface fluid. *Cell* **85**, 229–236 (1996).
- Lehrer, R. I., Lichtenstein, A. K. & Ganz, T. Defensins: antimicrobial and cytotoxic peptides of mammalian cells. *Annu. Rev. Immunol.* **11**, 105–128 (1993).
- Zabner, J., Smith, J. J., Karp, P. H., Widdicombe, J. H. & Welsh, M. J. Loss of CFTR chloride channels alters salt absorption by cystic fibrosis airway epithelia *in vitro*. *Mol. Cell* **2**, 397–403 (1998).
- Goldman, M. J. et al. Human β -defensin-1 is a salt-sensitive antibiotic in lung that is inactivated in cystic fibrosis. *Cell* **88**, 553–560 (1997).

22. Matsui, H. *et al.* Evidence for periciliary liquid layer depletion, not abnormal ion composition, in the pathogenesis of cystic fibrosis airways disease. *Cell* **95**, 1005–1015 (1998).
23. Tarran, R. *et al.* The CF salt controversy: *in vivo* observations and therapeutic approaches. *Mol. Cell* **8**, 149–158 (2001).
24. Knowles, M. R. *et al.* Ion composition of airway surface liquid of patients with cystic fibrosis as compared with normal and disease-control subjects. *J. Clin. Invest.* **100**, 2588–2595 (1997).
25. Selsted, M. E., Szklarek, D. & Lehrer, R. I. Purification and antibacterial activity of antimicrobial peptides of rabbit granulocytes. *Infect. Immun.* **45**, 150–154 (1984).
26. Song, Y., Salinas, D., Nielson, D. W. & Verkman, A. S. Hyperacidity of secreted fluid from submucosal glands in early cystic fibrosis. *Am. J. Physiol. Cell Physiol.* **290**, C741–C749 (2006).
27. Coakley, R. D. *et al.* Abnormal surface liquid pH regulation by cultured cystic fibrosis bronchial epithelium. *Proc. Natl Acad. Sci. USA* **100**, 16083–16088 (2003).
28. Laube, D. M., Yim, S., Ryan, L. K., Kisich, K. O. & Diamond, G. Antimicrobial peptides in the airway. *Curr. Top. Microbiol. Immunol.* **306**, 153–182 (2006).
29. Do, T. Q. *et al.* Lipids including cholesteryl linoleate and cholesteryl arachidonate contribute to the inherent antibacterial activity of human nasal fluid. *J. Immunol.* **181**, 4177–4187 (2008).
30. Singh, P. K., Tack, B. F., McCray, P. B. Jr & Welsh, M. J. Synergistic and additive killing by antimicrobial factors found in human airway surface liquid. *Am. J. Physiol. Lung Cell Mol. Physiol.* **279**, L799–L805 (2000).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank J. Bartlett, X. Chamling, J.-H. Chen, L. Durairaj, N. Gansemer, E. Hornick, P. Hughes, P. Ludwig, T. Mayhew, K. Mohn, L. Ostedgaard, M. Rector, L. Reznikov, L. Schneider, A. Shelton, T. Starner, P. Tan, A. Tucker, A. Walimbe and T. Yahr for assistance and/or discussion. This work was supported by the National Institutes of Health (NIH; HL51670, HL091842, HL102288) and the Cystic Fibrosis Foundation. D.A.S. was supported by AI076671 and Gilead Sciences Research Scholars Program in Cystic Fibrosis. H.P.H. was supported by Program Grant (RGP001612009-C) of the Human Frontier Science Program. M.J.W. is a Howard Hughes Medical Institute investigator.

Author Contributions A.A.P., X.X.T., M.J.H., M.H.A.A., S.R., D.A.S., P.B.M., M.J.W. and J.Z. designed studies. A.A.P., X.X.T., M.J.H., M.H.A.A., S.R., T.O.M., P.H.K., C.L.W.-L., H.P.H., M.v.E., B.B., A.R.H. and D.A.S. performed experiments. A.A.P., X.X.T., M.J.H., M.H.A.A., S.R., D.A.S., P.B.M., M.J.W. and J.Z. wrote the manuscript.

Author Information Microarray data have been deposited in the Gene Expression Omnibus under accession numbers GSE36906 and GSE21071. Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/nature. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to M.J.H. (michael-welsh@uiowa.edu) or J.Z. (joseph-zabner@uiowa.edu).

METHODS

CFTR^{-/-} and CFTR^{+/+} pigs. We previously reported generation of CFTR^{-/-} pigs^{7,31}. Animals were produced by mating CFTR^{+/+} male and female pigs. Newborn littermates were obtained from Exemplar Genetics. Animals were studied 8–15 h after birth. Euthanasia was with intravenous (i.v.) euthasol (Virbac). The University of Iowa Animal Care and Use Committee approved all animal studies.

Preparation of differentiated primary cultures of airway epithelia. Epithelial cells were isolated from the trachea and bronchi by enzymatic digestion, seeded onto collagen-coated, semi-permeable membranes (0.6 cm² Millicell-PCF; Millipore), and grown at the air–liquid interface as previously described^{12,32}. Culture medium, a 1:1 mixture of DMEM/F12, was supplemented with 2% Ultrosol G (PALL France SAS). Differentiated epithelia were used at least 14 days after seeding.

Bacteria-coated grids. Gold grids (TEM grids, 200 mesh, Ted Pella) were immersed in 1 mM 11-mercaptoundecanoic acid (HS(CH₂)₁₀COOH, MUA, Aldrich) solution for 30 min at room temperature, and then exposed to 1-ethyl-3-(3-diethylaminopropyl)carbodiimide (EDC) and N-hydroxysuccinimide (NHS) (molar ratio 1:2) for 30 min to activate the carboxyl groups of MUA. Grids were then placed in PBS, pH 7.4, containing 10 µg ml⁻¹ streptavidin (Sigma) for 30 min and rinsed with PBS. Streptavidin-coated grids were immersed in 1 M glycine for 30 min to quench the reaction and then rinsed with PBS. We used *S. aureus* isolate 43SA (isolated from a CF pig with pulmonary infection⁸), and *P. aeruginosa* PAO1. Molecular typing and antibiotic susceptibility profiling of *S. aureus* isolate 43SA were performed using standard methods. We determined that the 43SA strain is methicillin-sensitive and belongs to sequence type 398 (ST398), which is very common in pigs and often transferred to farmers³³. Bacteria were cultured to log-phase growth and 1×10^8 bacteria were mixed with 0.1 mg ml⁻¹ N-hydroxysulphosuccinimide (sulpho-NHS)-biotin (Thermo Scientific) for 1 h at room temperature. Streptavidin-coated grids were then incubated with biotinylated bacteria for 30 min and rinsed in PBS.

Bacteria-coated grids were used in three preparations. They were placed on the airway surface in pigs *in vivo*. They were tested in ASL removed from pigs after methacholine stimulation. And they were placed on the apical surface of primary cultures of differentiated airway epithelia. In each case, after removal from ASL, they were immediately (2–3 s) rinsed with PBS and then immersed in PBS containing the fluorescent indicators SYTO 9 and propidium iodide (Live/Dead BacLight Bacterial Viability assay, Invitrogen). Propidium iodide staining and this assay indicate the percentage of dead cells (Invitrogen, Molecular Probes product information and Supplementary Fig. 7). After 15 min, the grids were rinsed with PBS and placed on slides for imaging with a laser-scanning confocal microscope (Olympus FV1000). For each experiment, we tested bacteria-coated grids that were immersed in saline rather than exposed to ASL; the percentage of dead bacteria was 2–6%. In Fig. 1d, g, the shortest exposure to ASL was 30 s. When bacteria-coated grids were placed on the airway surface of wild-type pigs *in vivo* for a shorter time (5 s), $15 \pm 2\%$ ($n = 4$) of the bacteria were dead. Of note, a study using high-speed atomic force microscopy found that antimicrobial proteins disrupted the bacterial surface at variable rates ranging from ~10 s to minutes³⁴. During development of the bacteria-coated grid assay, we assessed other methods of labelling bacteria—including use of bacteria expressing green fluorescent protein and bacteria labelled with streptavidin-Alexa Fluor 647—other substrates including fibres and planar films, and other methods of attaching bacteria including antibody and protein A. The methods we adopted proved to be the most reliable for detecting live and dead bacteria.

For each experimental condition, time point and/or animal, two or three grids were examined. For each grid, all live and dead bacteria were counted in 5–16 individual microscopic ($\times 60$) fields; each field contained ~100–1,000 bacteria. The percentages of dead bacteria in each field were then averaged to determine the percentage of dead bacteria for the experimental data point. Operators were blinded to genotype. In some experiments, two different operators obtained similar results.

The total number of bacteria (live plus dead) attached to grids varied from experiment to experiment. For an individual grid, the number of bacteria per microscopic field also varied from field to field, with an average coefficient of variation of 0.56 for grids applied to ASL *in vivo* and *in vitro*. Supplementary Fig. 6 shows the average number of bacteria per microscopic field for the data in Fig. 1d, g. The total number of bacteria (live plus dead) in randomly selected microscopic fields did not differ between grids exposed to CF or non-CF ASL or in conditions in which many or a few bacteria were killed. In addition, the total number of bacteria per microscopic field (live plus dead, mean \pm s.d.) did not differ for bacteria exposed for 30 s to water (232 ± 94 total bacteria per field, 4.6% dead), saline (230 ± 103 total bacteria per field, 4.3% dead) or 70% alcohol (228 ± 108 total bacteria per field, 97% dead).

Bacteria-coated grids were also prepared for scanning electron microscopy using standard procedures. In brief, the grids were fixed in 2.5% glutaraldehyde in 0.1 M cacodylate buffer, followed by post-fixation in 1% osmium tetroxide. The grids were then dehydrated in a graded series of ethanol, transitioned to hexamethyldisilazane, air dried overnight and mounted on aluminium stubs. After sputter coating with gold/palladium, the samples were imaged in a Hitachi S-4800 scanning electron microscope.

Micro-CFU and radial diffusion assays. We used *S. aureus* strain 43SA and *E. coli* strain pCGLS-1. Micro-CFU assays were performed as previously described³⁵. When ASL was collected *in vivo*, each micro-CFU assay contained an initial inoculum of approximately 1×10^6 CFU ml⁻¹. When ASL was collected from cultured epithelia, each micro-CFU assay had an initial inoculum of approximately 3.3×10^3 CFU ml⁻¹. Radial diffusion assays with 10 mM sodium phosphate, pH 7.4, in 1% agarose were performed with a final attenuation ($D_{590\text{ nm}}$) of bacteria of 0.02 as previously described³⁶.

Collection of ASL for protein and antimicrobial studies. ASL was collected from pigs anaesthetized with ketamine (20 mg kg⁻¹, intramuscularly (i.m.)) and xylazine (2 mg kg⁻¹, i.m.), and maintained with propofol (2 mg kg⁻¹, i.v.). The neck was dissected to expose the trachea. Tracheal secretion was stimulated by administering methacholine (2.5 mg kg⁻¹, i.v.). After approximately 5 min, tracheal secretions were collected by making a small incision in the anterior tracheal wall and inserting a sterile polyester-tipped applicator (Puritan Medical Products) to collect ASL. The probe was then inserted into a microcentrifuge tube and secretions were recovered by centrifugation. This procedure produced approximately 10–20 µl of ASL fluid from each animal. For assays of ASL proteins, samples were immediately placed on ice and frozen at -80°C until use. For assays of antimicrobial activity, samples were used immediately.

ASL was also collected from primary cultures by rinsing the apical surface with 100 µl H₂O. On the basis of our earlier work²⁰, we estimate that collection produced an approximate 1:80 to 1:125 dilution of ASL.

Measurement of amounts of ASL antimicrobial proteins. To immunoblot for lactoferrin, PLUNC and SP-A, 10 µl of a 1:10 dilution of the ASL was separated on 4–15% Tris-HCl gels and transferred to polyvinylidene difluoride membranes, followed by blocking in TBS-Tween containing 2% BSA. Membranes were incubated with a primary antibody (rabbit anti-human lactoferrin, Immunology Consultants; monoclonal anti-human PLUNC, R&D Systems; or polyclonal antisera against porcine SP-A³⁷). Membranes were washed four times using TBS-Tween, then incubated with secondary antibody conjugated to horseradish peroxidase (Thermo Fisher Scientific) at a 1:20,000 dilution for 1 h. After five more washes in TBS-Tween, protein bands were detected using SuperSignal West Pico Chemiluminescent Substrate (Thermo Fisher Scientific). Membranes were then exposed to film and densitometry performed. The same ASL samples were used for each western blot and stripped between westerns with Restore Western Blot Stripping Buffer (Thermo Scientific).

Lysozyme was measured using lysoplates as previously described³⁶. Undiluted methacholine-stimulated secretions (5 µl) were used and compared with a standard of human lysozyme (Sigma).

Collection and determination of ASL Na⁺ and K⁺ concentrations. Pigs were anaesthetized with ketamine (20 mg kg⁻¹, i.m.) and xylazine (2 mg kg⁻¹, i.m.), followed by propofol (2 mg kg⁻¹, i.v.). The trachea was surgically exposed and accessed anteriorly using heat cautery. We then made a small anterior incision through the tracheal rings using heat cautery to prevent bleeding. To ensure that air was completely humidified, the animal was studied in a humidified chamber (100% relative humidity, 25–30 °C).

We collected ASL using a procedure designed to minimize the generation of excessive capillary forces during sampling. We fused thin lens paper (VWR Scientific Products) with Parafilm M (Pechiney Plastic Packaging) in an oven (205 °C) for 70–90 s. This procedure further reduced the volume of liquid that the paper would absorb and minimized evaporation from the surface not touching ASL. We prepared 0.5 \times 2 cm strips, washed them three times in double distilled water, and dried them overnight at 40 °C.

The Parafilm-fused paper strips were weighed and then gently placed in contact with the luminal surface of the posterior trachea for 15 s. Immediately after removal from the trachea, Parafilm-fused paper strips were placed (paper side up) on a precision balance (Mettler-Toledo XP26DR). Mass measurements were recorded by a synchronized computer ten times per second for 200 s (BalanceLink; Mettler-Toledo) while evaporation occurred. The amount of ASL collected was determined by plotting mass versus time, fitting a one-phase exponential decay to the data (GraphPad Prism 5; GraphPad Software), and extrapolating to mass at time 0, that is, the time at which the strip was removed from the airway surface. We then dried the strips overnight at 40 °C and measured the dry mass; 3.0 ± 0.5 µl of ASL ($n = 14$) was collected per cm².

To dissolve the dried ASL contents, the strips were placed in 1 ml flame photometer internal standard solution (Instrumentation Laboratory) overnight. We then measured Na^+ and K^+ content with an IL943 Flame Photometer and compared output with calibration curves of solutions containing known mole contents (Instrumentation Laboratory). ASL Na^+ and K^+ concentrations were calculated by dividing the mole content of cation by the mass of solvent. The mass of ASL solvent was determined as the difference between the initial ASL sample mass and the dry sample mass.

We also measured Na^+ and K^+ concentrations after methacholine administration (2.5 mg kg^{-1} , i.v.) to stimulate submucosal gland and goblet cell secretion³⁸. To collect the readily visible secretions, we lightly applied a sterile polystyrene applicator (Puritan Medical Products) to the tracheal surface. The applicator tip was then suspended and sealed in a microcentrifuge tube, and liquid was isolated by centrifugation through a layer of water-saturated oil. We diluted secretions 1:10 in double distilled H_2O , added $10 \mu\text{l}$ to 1 ml of flame photometer internal standard, measured Na^+ and K^+ content by flame photometry, and calculated the ion concentrations.

Measurement of ASL pH. To assess pH *in vivo*, we used non-invasive dual lifetime referencing to interrogate a $3 \times 3 \text{ mm}$ planar optode (pH sensitive foil, PreSens GmbH)³⁹ applied directly to the tracheal surface. The device used to transmit and receive the excitation and emission light was a single channel pH meter (pH-1 mini; PreSens GmbH). The tip of the fibre optic pH meter was kept at the same distance from the tracheal surface in all samples and confirmed by recording the amplitude registered by the device. Calibration before each set of measurements was done by placing the planar optode on the surface of a flat filter soaked in standard pH buffers. To minimize alterations in CO_2 during placement of the probe, the experiments were done in an environment of 5% CO_2 ; hence the CO_2 concentration in ASL was probably $>5\%$ owing to CO_2 production by the pigs. Thus, the pH values are probably lower than occur normally.

To assess pH *ex vivo*, secretion was stimulated with methacholine, and 5 min later, ASL was removed using the same methods as described for measurement of ion concentrations. Ten minutes after removal from the pig, ASL pH was measured using a needle-type fibre optic pH meter (World Precision Instruments). The pH meter was calibrated before each set of measurements. After removal and during measurement, the CO_2 over the sample was ambient, that is, approximately zero, which probably accounts for the higher pH values in the *ex vivo* ASL.

To assess pH in primary cultures of airway epithelia, we used the fluorescent ratiometric pH indicator SNARF conjugated to dextran (Molecular Probes). SNARF was prepared as a suspension in perfluorocarbon and $200 \mu\text{l}$ was added to the apical surface. Two hours later, epithelia were studied in a humidified, 5% CO_2 atmosphere at 37°C on the stage of an inverted laser-scanning microscope (Zeiss 510 Meta NLO). SNARF was excited at 488 nm, fluorescence intensity was measured at 580 nm and 640 nm, and pH was calculated as described⁴⁰.

Methods for changing CO_2 and for aerosolizing HCO_3^- . To decrease airway pH in non-CF pigs, CO_2 concentration was controlled in a humidified chamber (100% measured relative humidity, $25\text{--}30^\circ\text{C}$) using 0 or 15% CO_2 and 21% O_2 in N_2 . To increase airway pH in CF pigs, $50 \mu\text{l}$ of either 100 mM NaHCO_3 or 100 mM NaCl (as a control) were aerosolized (FMJ-250 and 1A-1C, Penn-Century) at 20 cm over the exposed trachea; that approach achieved an average deposition of $\sim 1 \mu\text{l cm}^{-2}$ of fluid.

Luminescence antibacterial assay. To assess the effect of pH on the antibacterial activity of lysozyme and lactoferrin, we used a luminescence assay of bacterial viability that has been described previously³⁰. The bacteria used were *S. aureus* Xen29 (Caliper LifeSciences Bioware) and *E. coli* DH5 α (GIBCO BRL, Life Technologies) expressing a luminescence plasmid pCGLS1. *E. coli* were grown in Luria-Bertani medium containing $100 \mu\text{g ml}^{-1}$ ampicillin (to maintain the plasmid) at 30°C with shaking. For *S. aureus*, $10 \mu\text{g ml}^{-1}$ kanamycin was used instead of ampicillin. Log-phase bacteria were centrifuged and resuspended in 1% Tryptic Soy Broth medium with 10 mM HEPES titrated to pH values from 6.8 to 8.0. *E. coli* (5×10^6 CFU) or *S. aureus* (5×10^5 CFU) were incubated with the indicated concentrations of lysozyme or lactoferrin in 96-well plates (Optiplate-96, Perkin Elmer) in a total volume of $120 \mu\text{l}$. Controls were the same conditions in the same plates, but without lysozyme or lactoferrin. After incubation at 30°C for the indicated times, luminescence was measured with a luminometer (Spectra Max L, Molecular Devices) and reported as RLU. A previous study⁴¹ determined that reductions in luminescence have an excellent correlation with a decrease in CFU.

Microarray analysis. Trachea and bronchus tissue samples were dissected from newborn piglets within 12 h of birth. Samples were cut into $\sim 5\text{-mm}^3$ pieces and stored in RNAlater RNA stabilization reagent (Ambion) using the manufacturer's recommended protocols. Primary cultures of differentiated CF and non-CF tracheal epithelia were prepared as described above. Total RNA was isolated with TRIzol reagent (Invitrogen). Only RNA samples attaining a minimum of 7.0 RNA integrity number on the Agilent 2100 Bioanalyzer (Agilent Technologies) were processed. Five micrograms of total RNA was used to generate biotinylated complementary RNA using the Affymetrix GeneChip one-cycle target labelling kit (Affymetrix) according to the manufacturer's recommended protocols, and then hybridized to the Affymetrix Porcine GeneChip (23,937 probe sets that interrogate approximately 23,256 transcripts from 20,201 *Sus scrofa* genes). Each hybridization sample (trachea, bronchus and cultured tracheal epithelia) was performed on a separate day, with all genotypes represented in each run. The arrays were washed, stained and scanned using the Affymetrix Model 450 Fluidics Station and Affymetrix Model 3000 scanner, and data were collected using the GeneChip Operating Software (GCOS), v.1.4, using the manufacturer's recommended protocols. Partek Genomics Suite Software (Partek) (one-way analysis of variance analysis) was used to analyse the data.

Quantitative RT-PCR. Primers to amplify porcine *GAPDH*, lactoferrin, lysozyme, S100A9 and PBD-2 were designed and validated using standard procedures. Total RNA was collected using TRIzol (manufacturer's recommended protocol) from the trachea and bronchus of 6 *CFTR*^{+/+} and 6 *CFTR*^{-/-} pigs and from primary cultures of tracheal epithelia from 8 *CFTR*^{+/+} and 8 *CFTR*^{-/-} pigs. Reverse transcription with $1 \mu\text{g}$ total RNA was performed using the high-capacity cDNA reverse transcription kit (Applied Biosystems). Ten nanograms of cDNA and 50 pm of both forward and reverse primers were used per reaction for quantitative PCR. Power SYBR green PCR master mix (Applied Biosystems) was used for quantification. *GAPDH* cDNA levels were used to normalize expression of lactoferrin, lysozyme, S100A9 and PBD-2. Fold change was generated using average $\Delta\Delta C_t$ values for each genotype. Error bars were generated using standard error of $\Delta\Delta C_t$ values of each sample. The primers used were: *GAPDH*, forward: 5'-GACTTCGAGCAGGAGA TGG-3', reverse: 5'-GCACGGTGTGGTGTGGCGTAGAGG-3'; lactoferrin, forward: 5'-AGCCATCGCTACCTGAAACATGC-3', reverse: 5'-ATCATGAAG GCACAGGCTTCCAG-3'; lysozyme, forward: 5'-TGCAAGAGGGGTGTGCA GAG-3', reverse: 5'-AAGAGACAAGGTGAGCTGAAG-3'; S100A9, forward: 5'-TCAGCCAGGCCCTATAAATGCTG-3', reverse: 5'-TCTTCCTGCATCT GTCCAAGC-3'; PBD-2, forward: 5'-GGATTGAAGGGACCTGTTACAG-3', reverse: 5'-GCAATACTTCACTTGGCCTG-3'.

Statistical analysis. Data are presented as mean \pm s.e.m. Unless otherwise indicated, statistical analysis used an unpaired Student's *t*-test. Differences were considered statistically significant at $P < 0.05$.

- Rogers, C. S. *et al.* Production of *CFTR*-null and *CFTR*-*ΔF508* heterozygous pigs by adenoassociated virus-mediated gene targeting and somatic cell nuclear transfer. *J. Clin. Invest.* **118**, 1571–1577 (2008).
- Karp, P. H. *et al.* in *Epithelial Cell Culture Protocols* Vol. 188 (ed. Wise, C.) 115–137 (Humana, 2002).
- Smith, T. C. *et al.* Methicillin-resistant *Staphylococcus aureus* (MRSA) strain ST398 is present in midwestern U.S. swine and swine workers. *PLoS ONE* **4**, e4258 (2009).
- Fantner, G. E., Barbero, R. J., Gray, D. S. & Belcher, A. M. Kinetics of antimicrobial peptide activity measured on individual bacterial cells using high-speed atomic force microscopy. *Nature Nanotechnol.* **5**, 280–285 (2010).
- Cole, A. M., Wu, M., Kim, Y. H. & Ganz, T. Microanalysis of antimicrobial properties of human fluids. *J. Microbiol. Methods* **41**, 135–143 (2000).
- Lehrer, R. I., Rosenman, M., Harwig, S. S. L., Jackson, R. & Eisenhauer, P. Ultrasensitive assays for endogenous antimicrobial polypeptides. *J. Immunol. Methods* **137**, 167–173 (1991).
- van Eijk, M. *et al.* Porcine pulmonary collectins show distinct interactions with influenza A viruses: role of the N-linked oligosaccharides in the carbohydrate recognition domain. *J. Immunol.* **171**, 1431–1440 (2003).
- Singer, M. *et al.* A MARCKS-related peptide blocks mucus hypersecretion in a mouse model of asthma. *Nature Med.* **10**, 193–196 (2004).
- Blossfeld, S. & Gansert, D. A novel non-invasive optical method for quantitative visualization of pH dynamics in the rhizosphere of plants. *Plant Cell Environ.* **30**, 176–186 (2007).
- Shuba, Y. M., Dietrich, C. J., Oermann, E., Cleemann, L. & Morad, M. Local extracellular acidification caused by Ca^{2+} -dependent exocytosis in PC12 cells. *Cell Calcium* **44**, 220–229 (2008).
- Travis, S. M. *et al.* Activity of abundant antimicrobials of the human airway. *Am. J. Respir. Cell Mol. Biol.* **20**, 872–879 (1999).

Dietary-fat-induced taurocholic acid promotes pathobiont expansion and colitis in *Il10*^{-/-} mice

Suzanne Devkota¹, Yunwei Wang¹, Mark W. Musch¹, Vanessa Leone¹, Hannah Fehlner-Peach¹, Anuradha Nadimpalli¹, Dionysios A. Antonopoulos², Bana Jabri¹ & Eugene B. Chang¹

The composite human microbiome of Western populations has probably changed over the past century, brought on by new environmental triggers that often have a negative impact on human health¹. Here we show that consumption of a diet high in saturated (milk-derived) fat, but not polyunsaturated (safflower oil) fat, changes the conditions for microbial assemblage and promotes the expansion of a low-abundance, sulphite-reducing pathobiont, *Bilophila wadsworthia*². This was associated with a pro-inflammatory T helper type 1 (T_H1) immune response and increased incidence of colitis in genetically susceptible *Il10*^{-/-}, but not wild-type mice. These effects are mediated by milk-derived-fat-promoted taurine conjugation of hepatic bile acids, which increases the availability of organic sulphur used by sulphite-reducing microorganisms like *B. wadsworthia*. When mice were fed a low-fat diet supplemented with taurocholic acid, but not with glycocholic acid, for example, a bloom of *B. wadsworthia* and development of colitis were observed in *Il10*^{-/-} mice. Together these data show that dietary fats, by promoting changes in host bile acid composition, can markedly alter conditions for gut microbial assemblage, resulting in dysbiosis that can perturb immune homeostasis. The data provide a plausible mechanistic basis by which Western-type diets high in certain

saturated fats might increase the prevalence of complex immune-mediated diseases like inflammatory bowel disease in genetically susceptible hosts.

Inflammatory bowel diseases (IBD) and other immune-related human disorders are relatively 'new' diseases in that their incidence has increased considerably over the past half century, matching developments in cultural westernization^{1,3,4}. The rapidity of these developments are probably not caused by genetic drift, but by exposure to non-genetic factors introduced through changes in the diet and lifestyle of genetically susceptible individuals, triggering aberrant host responses that lead to IBD. In this study, we examine whether certain dietary fats present in Western diets are capable of precipitating colonic inflammation through their actions on the enteric microbiota of genetically susceptible hosts.

The effects of three different diets (Supplementary Table 1) on the enteric microbiota of specific-pathogen-free (SPF) C57BL/6 mice are shown in Fig. 1a and Supplementary Table 2. With the exception of the low-fat purified mouse diet (LF), the high-fat diets were isocaloric and differed only in the type of dietary fat used, which was held constant at 37% of total calories and closely mimics Western consumption⁵. These fats also represent sources used in numerous processed

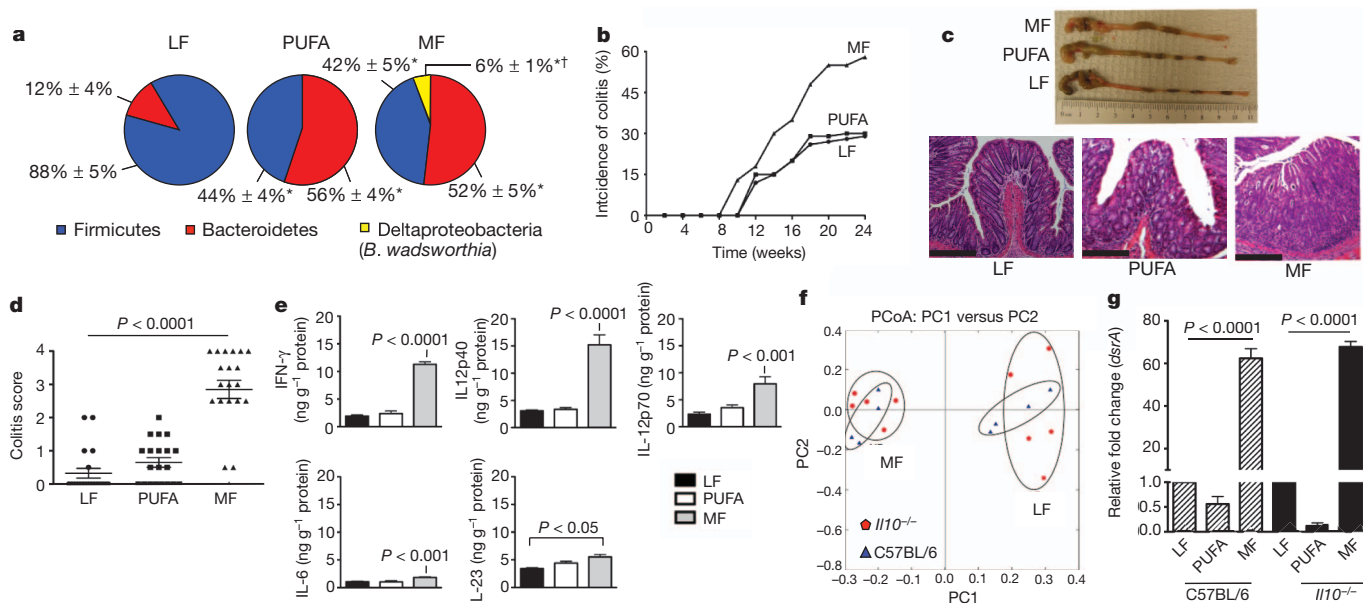


Figure 1 | Saturated MF-induced colitis is associated with bloom of *B. wadsworthia* in *Il10*^{-/-} mice. a–g. Samples from SPF C57BL/6 (a, g) (*n* = 6 per group) and SPF *Il10*^{-/-} mice (b–g) fed MF, PUFA or LF for 24 weeks (*n* = 20 per group). a, Phyla representation shown for LF, PUFA and MF with means ± s.e.m. **P* < 0.05 compared to LF, †*P* < 0.05 compared to PUFA and LF. b, Gross incidence of colitis. c, Representative colon lengths (top) and

haematoxylin and eosin staining of distal colon (bottom). Scale bars, 400 μm. d, Blinded histological colitis scores²⁵. e, Distal colonic mucosal cytokines determined by ELISA. f, Principal coordinates analysis (PCoA) plot of the UNIFRAC metric matrix. PC, principal coordinate. g, qPCR of caecal content *dsrA* (normalized to LF diet).

¹Department of Medicine, Section of Gastroenterology, The University of Chicago, Knapp Center for Biomedical Discovery, 900 East 57th Street, Chicago, Illinois 60637, USA. ²Institute for Genomics and Systems Biology, Argonne National Laboratory, 9700 South Cass Avenue, Argonne, Illinois 60439, USA.

and confectionary foods. Twenty-one-day exposure to the three study diets resulted in significant differences in the structure of the enteric microbiota as assessed by both Sanger-based and 454-based DNA sequencing of 16S ribosomal RNA (rRNA) libraries from caecal contents and stool. Both high-fat diets reduced the richness of the microbiota compared with LF (Supplementary Fig. 1). LF promoted Firmicutes, but also resulted in a lower abundance of most other phyla, whereas polyunsaturated (safflower oil) fat (PUFA) and saturated (milk-derived) fat diets (MF) resulted in a higher abundance of Bacteroidetes and a lower abundance of Firmicutes. Interestingly, these changes differed from those induced by lard-based, saturated fats^{6,7}

(Supplementary Fig. 2). Whereas MF and PUFA had similar effects on Bacteroidetes and Firmicutes, a significant bloom of a member of the Deltaproteobacteria, *B. wadsworthia*, was consistently observed only with MF. *B. wadsworthia* is a sulphite-reducing, immunogenic microbe that is difficult to detect in healthy individuals, but emerges under pathological conditions such as appendicitis and other intestinal inflammatory disorders^{8–16}.

MF did not affect wild-type mice, but increased the onset and incidence of colitis in *Il10*^{-/-} mice, driving it from a spontaneous rate of 25–30% (on LF) to over 60% in a 6-month period (Fig. 1b). In contrast, the incidence of colitis in *Il10*^{-/-} mice fed PUFA was no

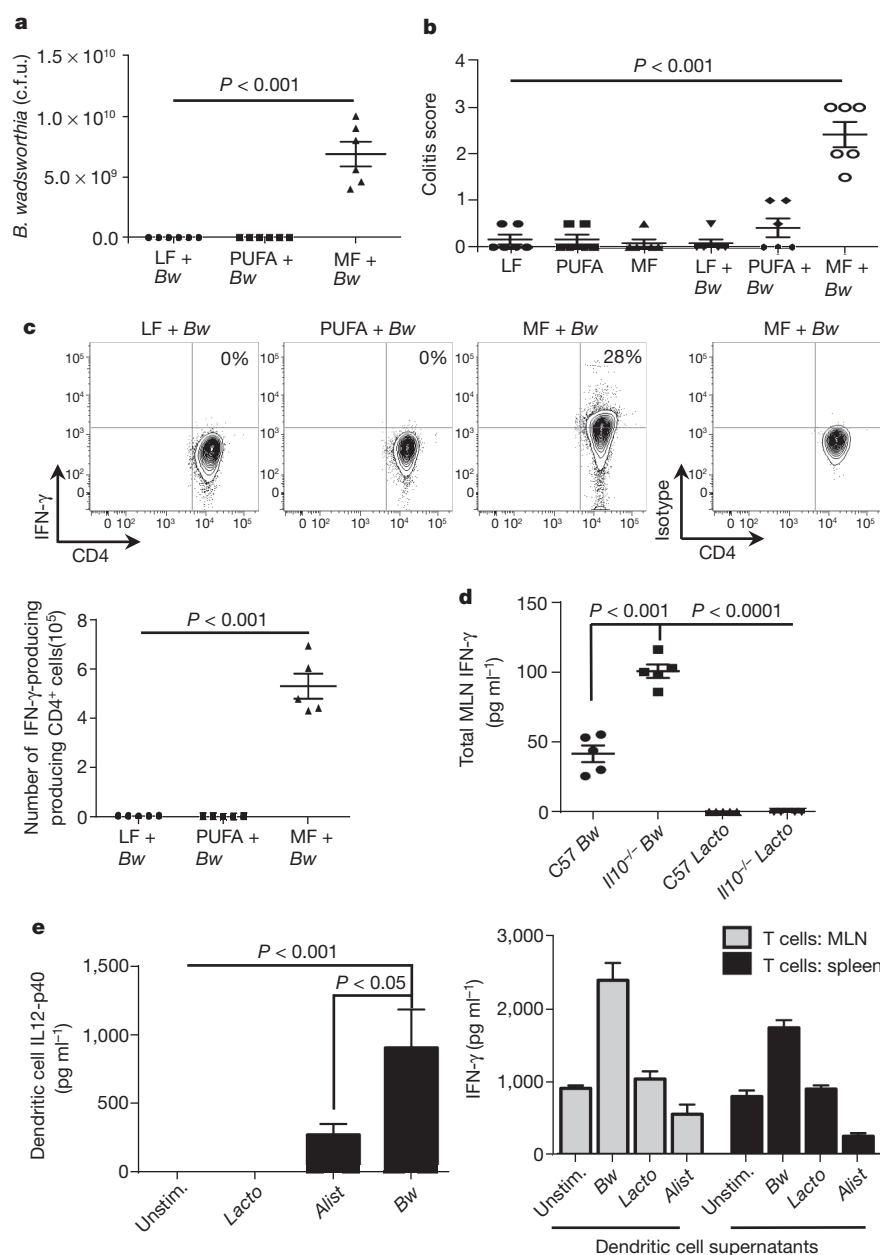


Figure 2 | *B. wadsworthia* mono-association in GF *Il10*^{-/-} mice can only be established with consumption of MF diet, resulting in a T_H1 immune response and development of colitis. **a–c**, Samples from GF *Il10*^{-/-} mice \pm mono-association with 10^8 c.f.u. *B. wadsworthia* (Bw) maintained on LF, PUFA or MF for 5 weeks ($n = 5$ per group). **a**, c.f.u. counts of cultured caecal-derived *B. wadsworthia*. **b**, Blinded histological colitis scores. **c**, IFN- γ production by CD4⁺ T cells in MLNs. **d**, IFN- γ production by MLNs from GF C57BL/6 and *Il10*^{-/-} mice colonized with either *B. wadsworthia* or *L. murinus*

(Lacto) and restimulated *ex vivo* with pure culture lysate from the respective bacterium. **e**, *In vitro* CD4⁺ T-cell differentiation assay. Left, IL-12p40 produced by dendritic cells challenged with pure lysates from *B. wadsworthia*, *L. murinus* or *Alistipes* (Alist). Unstim., unstimulated. Data represent pooled values from MLN dendritic cells and splenic dendritic cells in the presence of retinoic acid/TGF- β . Right, IFN- γ production by CD4⁺ T cells stimulated with supernatants from the bacteria-challenged dendritic cells. Data shown represent one out of two assays, performed in triplicate.

different than those fed LF. The colitis seen in mice fed MF was also more severe and extensive (Fig. 1c). These changes paralleled differences in histological colitis scores (Fig. 1d). Inflammatory mucosal cytokine levels from the distal colon were significantly raised compared to LF and in most cases, PUFA (Fig. 1e). LF, PUFA and MF elicited effects on the total enteric microbiota of *Il10*^{-/-} mice that were similar to those observed in wild-type mice (shown for LF and MF, Fig. 1f). Similarly, *B. wadsworthia*, as detected by quantitative polymerase chain reaction (qPCR) of the dissimilatory sulphite reductase A gene unique to sulphite-reducing bacteria of which *B. wadsworthia* is the most prominent in our model, was found at equal relative abundance in mice on MF, independent of genotype (Fig. 1g). The bloom of *B. wadsworthia* induced by MF was also observed in dextran sodium sulphate (DSS)-treated SPF C57BL/6 mice, in which the onset and severity of colitis were more severe than that seen in LF- and PUFA-fed mice (Supplementary Fig. 3). Altogether these observations suggest that the bloom of sulphite-reducing Deltaproteobacteria, particularly *B. wadsworthia*, is associated with colitis in hosts that are genetically susceptible or have compromised mucosal barrier function.

To explore whether a MF diet was necessary for the survival and proliferation of *B. wadsworthia*, we mono-associated germ-free (GF) *Il10*^{-/-} mice with *B. wadsworthia* that were consuming either LF, PUFA or MF. Five weeks after gavage, colonization of the colon could only be established in mice fed MF (Supplementary Fig. 4, shown for LF and MF), whereas on LF, *B. wadsworthia* was undetectable. *B. wadsworthia* identity was confirmed by PCR of the 16S rRNA encoding gene from caecal-derived DNA using universal primers, followed by direct sequencing of the PCR product and cultivation of *B. wadsworthia* from caecal contents (Fig. 2a). Additionally, qPCR analysis of luminal- versus mucosal-associated *B. wadsworthia* revealed a nearly 45-fold increase in the latter (Supplementary Fig. 5).

MF, in contrast to LF or PUFA, increased the incidence of colitis (Fig. 2b), although to a lesser extent than that seen in SPF *Il10*^{-/-} mice (Supplementary Fig. 6a). Furthermore, increased levels of interferon- γ (IFN- γ), interleukin-12 p40 subunit (IL-12p40) and interleukin-12 p70 heterodimer (IL-12p70), as well as low or undetectable levels of interleukin-6 (IL-6), IL-17 and IL-23 in the colonic mucosa of these mice, were consistent with the induction of a distinct T_H1 immune response (Supplementary Fig. 6b). This was further confirmed by increased CD4⁺ IFN- γ ⁺ populations in the mesenteric lymph nodes (MLNs) of mice colonized on MF (Fig. 2c). These changes were not observed in mice consuming MF in the absence of *B. wadsworthia*, indicating that the diet itself is not immunogenic. Moreover, when *Lactobacillus murinus*, which is also promoted by MF, was mono-associated in GF *Il10*^{-/-} mice, no evidence of colitis or immune activation was seen (data not shown). The specificity of the T_H1 response induced by *B. wadsworthia* was further elucidated by both *ex vivo* and *in vitro* challenges of immune cells using pure bacterial lysates from *B. wadsworthia*, *L. murinus* and *Alistipes*, the latter two representing bacteria that were promoted by MF (albeit to a lesser degree, Supplementary Table 2). Only lysate from *B. wadsworthia* elicited an IFN- γ response in MLNs harvested from *B. wadsworthia* mono-associated GF *Il10*^{-/-} mice (Fig. 2d). The response in *Il10*^{-/-} MLNs was much greater than that in C57BL/6 MLNs, probably due to the absence of IL-10 modulating signals. An *in vitro* T-cell differentiation assay was then performed in which MLN and splenic dendritic cells were isolated from SPF C57BL/6 mice and stimulated with pure lysates from *B. wadsworthia*, *L. murinus* or *Alistipes* (plus retinoic acid and TGF- β for splenic dendritic cells). Supernatants from these dendritic cells revealed significantly elevated levels of IL-12p40 only in the *B. wadsworthia* stimulated dendritic cells (Fig. 2e, left). When purified, T cells were then incubated for 3 days with dendritic cell supernatants

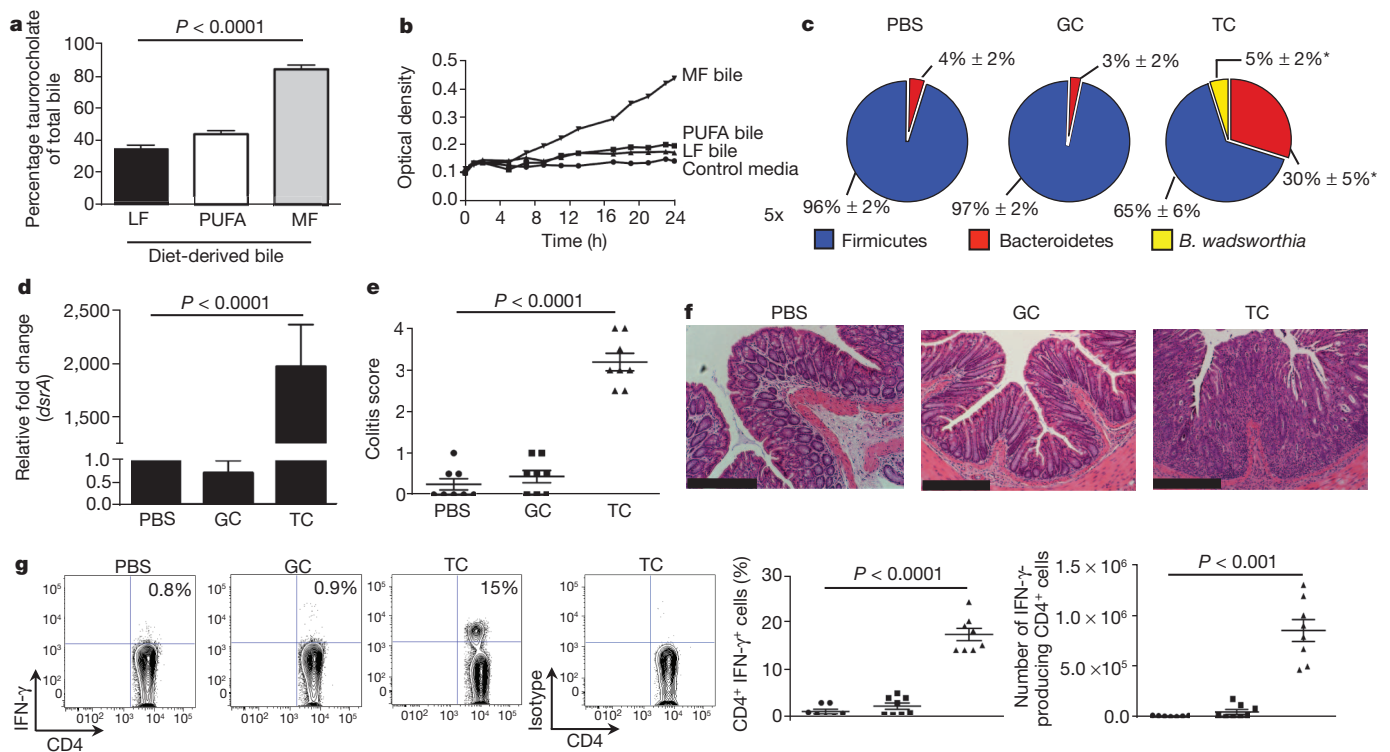


Figure 3 | Induction of TC bile acid following consumption of MF promotes bloom of *B. wadsworthia* both *in vitro* and in SPF *Il10*^{-/-} mice, resulting in colitis. **a**, TC content of gall bladder aspirates from *Il10*^{-/-} mice consuming LF, PUFA or MF for 5 weeks. **b**, Growth curve of *B. wadsworthia* in media containing gall bladder aspirates. **c–g**, Samples from SPF *Il10*^{-/-} mice gavaged with PBS, TC or GC daily for 21 days while maintained on LF diet

($n = 8$ per group). **c**, Phyla representation with bloom of *B. wadsworthia* in the TC group with means \pm s.e.m. * $P < 0.05$ compared to PBS and GC. **d**, Relative abundance of *dsrA* in caecal contents (by qPCR and normalized to LF diet). **e, f**, Blinded histological colitis scores (**e**) and haematoxylin and eosin staining of distal colon (**f**). Scale bars, 400 μ m **g**, IFN- γ production in MLNs.

from each respective treatment and analysed for IFN- γ and IL-17. MLN and splenic T cells incubated with supernatants from dendritic cells exposed to *B. wadsworthia* produced nearly twofold higher levels of IFN- γ compared to T cells incubated with supernatants of dendritic cells exposed to *Alistipes* or *L. murinus* (Fig. 2e, right). In accordance with *in vivo* data, T cells did not produce IL-17. Therefore, *B. wadsworthia* seems to activate dendritic cells in a way that selectively induces a T_H1-mediated colitis. Of note, the possibility that other *B. wadsworthia* by-products (for example, H₂S, secondary bile acids) also stimulate dendritic cells cannot be excluded.

B. wadsworthia flourishes in the presence of taurine-conjugated (TC) bile acid (a property from which it got its name), a rich source of organic sulphur, which is used as the terminal electron acceptor of the electron transport chain resulting in the formation of H₂S as a by-product¹⁷. Because of their hydrophobicity, milk fats will promote increased hepatic taurine conjugation of bile acids, which are more efficient for micelle formation and fat emulsification^{18–20}. This was confirmed by mass spectrometry measurements of gall bladder aspirates from mice fed LF, PUFA and MF (Fig. 3a). When 10⁷ colony-forming units (c.f.u.) of *B. wadsworthia* in taurine-free liquid growth media were supplemented with 20 μ l of gall bladder aspirates obtained from SPF C57BL/6 mice fed the three test diets ($n = 5$ pooled), *B. wadsworthia* growth was selectively and robustly stimulated only by bile from MF-fed mice (Fig. 3b). To determine whether the dietary effect was in fact mediated by TC, SPF *Il10*^{−/−} mice were fed LF and gavaged with either TC or glycocholic acid (GC) daily for 3 weeks. This resulted in a bloom of *B. wadsworthia* with TC (Fig. 3c), nearly identical to that which was observed with consumption of MF.

In contrast, GC and phosphate-buffered saline (PBS) had little effect and *B. wadsworthia* remained undetectable (Fig. 3d). The bloom of *B. wadsworthia* observed in TC-gavaged mice was associated with increased incidence and severity of colitis (Fig. 3e, f). T_H1 cytokines were increased in both the mucosa (Supplementary Fig. 7) and MLNs (Fig. 3g). In further support of the role of bile in MF-diet-induced pathogenesis, mono-association with *B. wadsworthia* can be established in GF *Il10*^{−/−} mice when accompanied by TC administration, but not by GC or PBS (Fig. 4a), as demonstrated by the re-isolation of *B. wadsworthia* from the caecal contents of TC-fed mice (note that black colour change indicates H₂S production, Fig. 4a, left), and confirmed by colony counts (Fig. 4a, right). These mice developed colitis (Fig. 4b and Supplementary Fig. 8a) and again exhibited elevated T_H1 mucosal responses (Fig. 4c and Supplementary Fig. 8b). Together these findings indicate that the bloom of *B. wadsworthia* promoted by these dietary factors is selectively associated with T_H1 immunity.

We find the dependence of *B. wadsworthia* on diet-induced taurocholic acid intriguing and possibly representative of how certain gut microbes use bile to their advantage. Bile formation is unique to vertebrates, providing the host with the ability to digest and utilize a far greater variety of dietary substrates. Bile also has potent antimicrobial properties that can contribute to the selection or exclusion of many potential gut microbiota. However, several intestinal pathogens, including protozoa such as *Giardia*, Microsporidia and *Cryptosporidia*, and bacteria such as *B. wadsworthia*, *H. hepaticus* and *Listeria monocytogenes*, are not only bile-resistant, but highly favoured in the presence of bile^{21,22}, possibly through suppression of symbiotic, commensal microorganisms, allowing pathobionts and pathogens an opportunity to

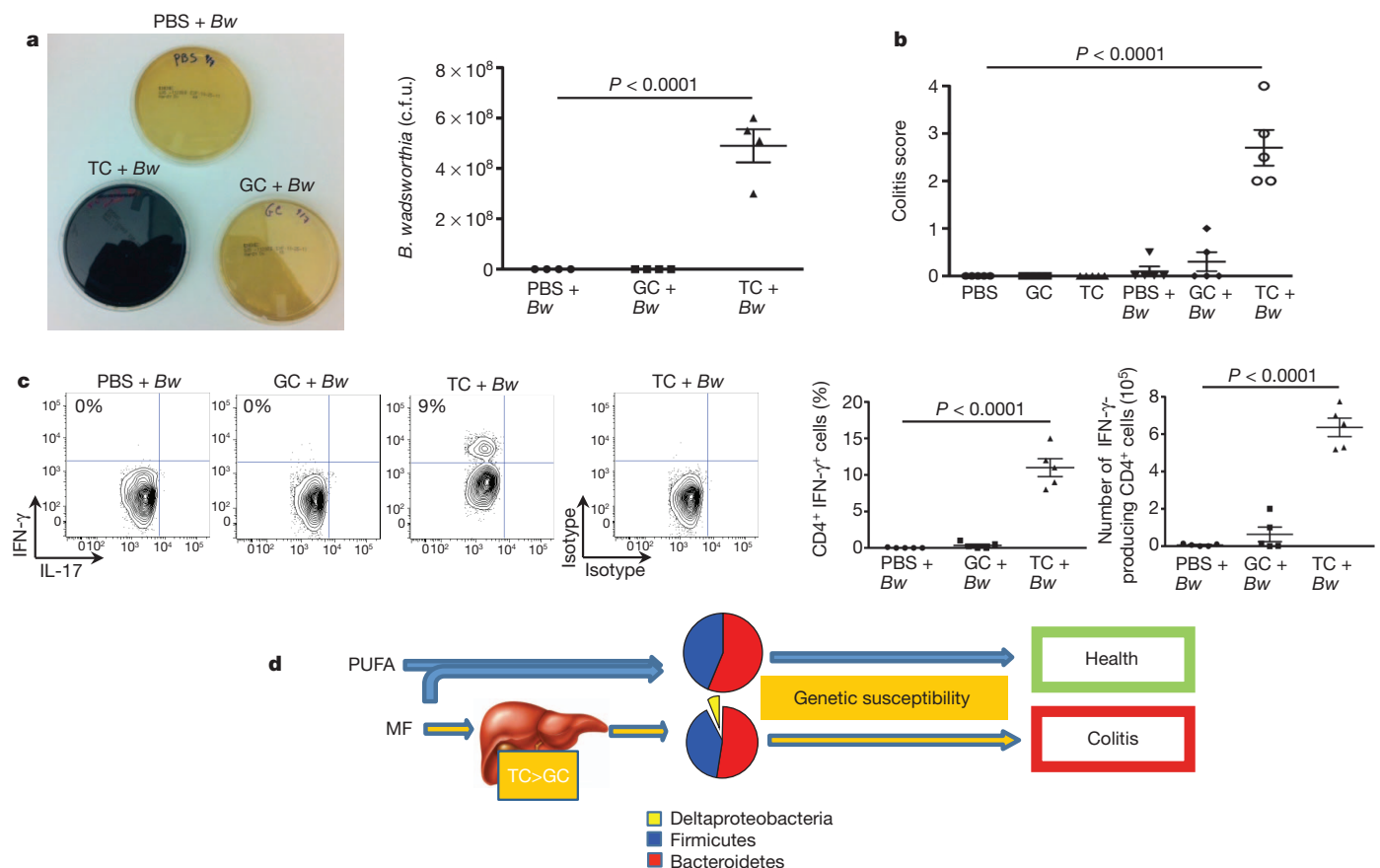


Figure 4 | Mono-association with *B. wadsworthia* in GF *Il10*^{−/−} mice is successful only if accompanied by TC gavage. **a–c**, Samples from GF *Il10*^{−/−} mice fed LF \pm mono-association with *B. wadsworthia* (Bw) followed by daily gavage with PBS, GC or TC for 21 days ($n = 5$ per group). **a**, Robust *B. wadsworthia* growth when re-isolated from caecal content of TC-fed GF mice

(black film in TC plate indicates H₂S production), and c.f.u. counts of caecal-derived *B. wadsworthia*. **b**, Blinded histological colitis scores. **c**, IFN- γ production in MLN CD4⁺ T cells determined by intracellular staining. **d**, Proposed experimental model.

establish a niche in the intestine. Once established, the by-products of these bacteria, whether H₂S or secondary bile acids, can serve as gut mucosal 'barrier breakers', allowing for increased immune-cell infiltration and thus acting synergistically with the bacterial antigen-specific immune response to induce tissue damage. In genetically susceptible hosts, this development has the capacity to tip a compensated state of immune balance in favour of chronic disease.

METHODS SUMMARY

Extraction of bacterial DNA, clone library preparation and sequencing were conducted as previously described^{23,24}. Amplicon libraries were also used to target the V3–V4 region of the 16S rRNA encoding gene for deeper sequence surveys of the microbial communities. *B. wadsworthia* cultures were maintained in liquid and solid growth media as previously described⁸. GF mice were mono-associated by oral gavage with 10⁸ c.f.u. at a single time point. Bile acids were orally gavaged at 1 g kg⁻¹ body weight in 100 µl PBS. Inoculation of *B. wadsworthia* liquid cultures with bile aspirate was performed by collecting gall bladder contents from mice adapted to MF, PUFA or LF and anaerobically adding them to *B. wadsworthia* in taurine-free media. Optical density (OD) was then measured every hour for 24 h. Re-isolation of *B. wadsworthia* from host was achieved by streaking 1:10 serial dilutions of 5 g caecal contents on BBE agar plates, which were allowed to grow anaerobically for 4 days. Histological colitis indices were scored blindly as previously described²⁵. Mucosal cytokines were measured by ELISA using 50 µg protein derived from scraping the length of the distal colon. Intracellular cytokines were measured in cells isolated from MLNs and stimulated for 3 h with PMA and ionomycin. qPCR used primers donated by J. Tiedje, targeting the *dsrA* gene. For *in vitro* immune assays, dendritic cells were isolated from MLNs and spleens of C57BL/6 mice and stimulated with bacteria lysates for 24 h in the presence of retinoic acid and TGF-β (for spleens) and measured for IL-12p40. Supernatants from these cultures were used to challenge naive T cells isolated from MLNs and spleens of C57BL/6 mice for 3 days and measured for IFN-γ and IL-17. Statistical analyses were performed using one-way ANOVA with Tukey post-hoc.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 19 October 2011; accepted 10 May 2012.

Published online 13 June 2012.

- Walter, J. & Ley, R. The human gut microbiome: ecology and recent evolutionary changes. *Annu. Rev. Microbiol.* **65**, 411–429 (2011).
- Baron, E. J. *Bilophila wadsworthia*: a unique Gram-negative anaerobic rod. *Anaerobe* **3**, 83–86 (1997).
- Molodecky N. A. *et al.* Increasing incidence and prevalence of the inflammatory bowel diseases with time, based on systematic review. *Gastroenterology* **142**, 46–54 (2012).
- Maslowski, K. M. & Mackay, C. R. Diet, gut microbiota and immune responses. *Nature Immunol.* **12**, 5–9 (2011).
- NHANES. Trends in intake of energy and macronutrients—United States 1971–2000 <http://www.cdc.gov/mmwr/preview/mmwrhtml/mm5304a3.htm> (2004).
- Turnbaugh, P., Backhed, F., Fulton, L. & Gordon, J. I. Diet-induced obesity is linked to marked but reversible alterations in the mouse distal gut microbiome. *Cell Host Microbe* **3**, 213–223 (2008).
- Hildebrandt, M. A. *et al.* High-fat diet determines the composition of the murine gut microbiome independently of obesity. *Gastroenterology* **137**, 1716–1724 (2009).
- Baron, E. J. *et al.* *Bilophila wadsworthia*, a unique Gram-negative anaerobic rod recovered from appendicitis specimens and human faeces. *J. Gen. Microbiol.* **135**, 3405–3411 (1989).
- Attene-Ramos, M. S., Wagner, E. D., Plewa, M. J. & Gaskins, H. R. Evidence that hydrogen sulfide is a genotoxic agent. *Mol. Cancer Res.* **4**, 9–14 (2006).

- Loubinoux, J., Bronowicz, J.-P., Peireira, I. A. C., Mougnet, J. L. & Le Faou, A. E. Sulphate reducing bacteria in human faeces and their association with inflammatory diseases. *FEMS Microbiol. Ecol.* **40**, 107–112 (2002).
- Rowan, F. E., Docherty, N. G., Coffey, J. C. & O'Connell, P. R. Sulphate-reducing bacteria and hydrogen sulphide in the etiology of ulcerative colitis. *Br. J. Surg.* **96**, 151–158 (2009).
- Beech, I. B. & Zinkevich, V. Screening of sulfate-reducing bacteria in colonoscopy samples from healthy and colitic human gut mucosa. *FEMS Microbiol. Ecol.* **2**, 147–155 (2000).
- Gibson, G. R., Cummings, J. H. & Macfarlane, G. T. Growth and activities of sulphate-reducing bacteria in gut contents of healthy subjects and patients with ulcerative colitis. *FEMS Microbiol. Ecol.* **86**, 103–112 (1991).
- Scanlan, P. D., Shanahan, F. & Marchesi, J. Culture-independent analysis of desulfovibrios in the human distal colon of healthy, colorectal cancer and polypectomized individuals. *FEMS Microbiol. Ecol.* **69**, 213–221 (2009).
- Arzese, A., Mercuri, F., Trevisan, R., Menozzi, G. & Botta, G. Recovery of *Bilophila wadsworthia* from clinical specimens in Italy. *Anaerobe* **3**, 219–224 (1997).
- Baron, E. J. *et al.* *Bilophila wadsworthia* isolates from clinical specimens. *J. Clin. Microbiol.* **30**, 1882–1887 (1992).
- Laue, H., Denger, K. & Cook, A. M. Taurine reduction in anaerobic respiration of *Bilophila wadsworthia* RZATAU. *Appl. Environ. Microbiol.* **63**, 2016–2021 (1997).
- Lindstedt, S., Avigan, J., Goodman, D. S., Sjoval, J. & Steinberg, D. The effects of dietary fat on the turnover of cholic acid and on the composition of the biliary bile acids in man. *J. Clin. Invest.* **44**, 1754–1765 (1965).
- Rueda, A. *et al.* Conjugated bile acids and intestinal flora during the pre-ruminant stage in goat: influence of a lamb milk replacer. *Arch. Physiol. Biochem.* **104**, 246–251 (1996).
- Graham, T. O., Van Thiel, D. H., Little, J. M. & Lester, R. Synthesis of taurocholate by rat fetal liver in organ culture: effects of cortisol *in vitro*. *Am. J. Physiol.* **237**, E177–E184 (1979).
- Ananieva, O., Nilsson, I., Vorobjova, T., Uibo, R. & Wadström, T. Immune responses to bile-tolerant *Helicobacter* species in patients with chronic liver diseases, a randomized population group, and healthy blood donors. *Clin. Vaccine Immunol.* **9**, 1160–1164 (2002).
- Dussurget, O., Cabanes, D., Dehoux, P. & Lecuit, M. *Listeria monocytogenes* bile salt hydrolase is a PrfA-regulated virulence factor involved in the intestinal and hepatic phases of listeriosis. *Mol. Microbiol.* **4**, 1098–1106 (2002).
- Wang, Y. *et al.* Regional mucosa-associated microbiota determine physiological expression of TLR2 and TLR4 in murine colon. *PLoS ONE* **5**, e13607 (2010).
- Wang, Y. *et al.* 16S rRNA gene-based analysis of fecal microbiota from preterm infants with and without necrotizing enterocolitis. *ISME J.* **3**, 944–954 (2009).
- Berg, D. J. *et al.* Enterocolitis and colon cancer in interleukin-10-deficient mice are associated with aberrant cytokine production and CD4⁺ T_H1-like responses. *J. Clin. Invest.* **4**, 1010–1020 (1996).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by the National Center for Research Resources and the NIDDK, NIGMS and NCCAM of the National Institutes of Health through grant number DK-42086 (E.B.C.), DK47722 (E.B.C.), UH3DK083993 (E.B.C.), F31AT006073 (S.D.). Also, the Gastrointestinal Research Foundation, Crohns and Colitis Foundation of America (Y.W.), the Peter and Carol Goldman Family Research Fund, and the Harry and Leona Helmsley Trust Foundation (SHARE). We are also indebted to S. Finegold for his suggestions on successful culture of *B. wadsworthia* and *Alistipes*, J. Tiedje and M. Vital for *dsrA* primer sequences, L. Hagey for mass spectrometry, E. Huang, B. Theriault and J. Stencel for assistance with experiments, and R. Bouziat for T-cell purification.

Author Contributions S.D. and E.B.C. were involved in all aspects of this study, especially in the development of the hypothesis, experimental plan and data analysis. Y.W., M.W.M., V.L., H.F.-P. and A.N. helped perform the experiments. D.A.A. and B.J. provided critical feedback and expertise and assisted in the analysis of data.

Author Information Data have been deposited in GenBank under accession numbers JQ890637–JQ894320. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to E.B.C. (echang@medicine.bsd.uchicago.edu).

METHODS

Mice. All mice were bred in-house and housed in our SPF and GF animal facilities. GF mice were maintained on 18% (by weight) 37% (kcal) MF, lard or PUFA test diets formulated by Harlan-Teklad and irradiated and tested for sterility both before and after use. LF purified diet was based on AIN-93M. All bacterial gavages (*B. wadsworthia*, ATCC 49260; *Alistipes*, ATCC BAA-1179; *L. murinus*, ATCC 35020) were performed using 10^8 c.f.u., and taurocholate and glycocholate (Sigma-Aldrich) using 1 g kg^{-1} dissolved in $100 \mu\text{g}$ PBS. Large particles of glycocholate were first filtered out before gavaging. All experiments were performed in accordance with the Institutional Biosafety Committee and the Institutional Care and Use Committee.

DSS colitis. SPF C57BL/6 mice were fed LF, PUFA or MF for 1 week ($n = 5$ per group). On day 7, 1.5% DSS was added to drinking water. On day 12, mice were changed back to plain tap water, and on day 15 mice were terminated. Weight loss and stool consistency was checked daily.

Clone library sequencing. Raw sequence data was processed using the Ribosomal Database Project's Pipeline Tool (<http://rdp.cme.msu.edu/>), which includes base calling, quality trimming and alignment as part of the workflow. Additionally, chimaeric sequences were screened and removed using Bellerophon. The RDP Classifier was used to assign the 16S rRNA sequences to a hierarchical taxonomy. The program mothur was used to group sequences into operational taxonomical units (OTUs) using the furthest neighbour algorithm and a 97% sequence similarity criterion. For PCoA, all 16S rRNA gene sequences were imported into the ARB software package and aligned into a phylogenetic tree by neighbour joining, which was used to perform clustering analysis using online UniFrac without abundance weighting.

16S rRNA-based amplicon library preparation and data analysis. PCR primers used were specific for the V3–V4 region of the 16S rRNA encoding gene (*Escherichia coli* positions 338–802; 338F: 5'-ACTCCTACGGGAGGCAGC-3'; and equimolar amounts of 802R-A 5'-TACCRGGGTHCTAATCC-3', 802R-B 5'-TACCAGAGTATCTAATTC-3', 802R-C 5'-CTACDSRGGTMTCTAATC-3', 802R-D 5'-TACNVGGGTATCTAATCC-3', and contained 454-specific adaptor sequences as well as an 8-base-pair barcode. This barcode-based primer approach allowed sequencing of multiple samples in a single 454 sequencing run without the need for physical partitioning. Sequencing was performed at the High-Throughput Genome Analysis Core (HGAC; part of the Institute for Genomics and Systems Biology) at Argonne National Laboratory. Sequences were then trimmed and classified with the QIIME toolkit. Using the QIIME wrappers, OTUs were picked at 97% sequence identity using cdhit and a representative sequence was then chosen for each OTU by selecting the most abundant sequence in that OTU. These representative sequences were aligned using PyNAST and taxonomy was assigned to them using the RDP Classifier. The PyNAST-aligned sequences were also used to build a phylogenetic tree with FastTree and unweighted UniFrac distances then computed between all samples for additional ecological analyses, including PCoA.

Data are available to the public via the MG-RAST system (<http://metagenomics.anl.gov/>), including instant availability of the sequence data, bioinformatic analyses and tools, plus the support for export to QIIME.

Sulphite-reducing bacteria quantification. Sulphite-reducing bacteria were quantified using specific primers developed by M. Vital and J. Tiedje (MSU) for the *dsrA* gene. Forward, 5'-CCAACATGCACGGYTCCA-3'; reverse, 5'-CGTC GAACCTGAACTTGAACCTGTAGG-3'.

Bile composition analysis. Bile acid conjugates in pooled gall bladder aspirates were analysed by HPLC as described previously²⁶. Conjugated bile acids were quantified in the column effluent by monitoring the absorbance at 205 nm (for the amide bond). Peaks were identified using the relative retention time of known standards. Next, electrospray mass spectrometry was performed on the pooled gall bladder aspirates using a Perkin-Elmer Sciex API-III instrument modified with a nanoelectrospray source. The instrument was operated in the negative mode with Q1 IS voltage set to 600 V. The IN and ORI voltages were set to 110 V and 90 V respectively. Chemical identity of peaks was confirmed by the fragmentation pattern of selected ions (Q3 mode) using argon collision gas. The presence of conjugated bile salts was confirmed by selection for m/z 74 (glycine), m/z 97 (sulphate), m/z 124 (taurine).

T-cell and dendritic-cell purification. For CD4⁺ T-cell isolation, spleens and MLNs were mechanically disrupted through a 70- μm cell strainer. CD4⁺ cells were isolated first by CD25 negative selection using anti-CD25 APC with anti-APC microbeads on automacs. This was followed by positive immunoselection using CD4-(L3T4) microbeads (Miltenyi Biotec).

For dendritic-cell isolation, MLN and spleen were digested with 400 units ml^{-1} collagenase type IV (Sigma-Aldrich). Cells were filtered, resuspended in 22.5% Optiprep (Sigma-Aldrich), overlaid with Hank's buffered saline (HBS) and centrifuged at 670g for 30 min. Dendritic cells were then enriched from the interface by positive immunomagnetic selection using anti-CD11c-coated beads according to the manufacturer's recommendations (Miltenyi Biotec). Purification yielded up to 90% CD11c⁺ cells.

In vitro T-cell differentiation assay. Dendritic cells from MLNs and spleen plus retinoic acid/TGF- β were incubated for 24 h with $25 \mu\text{g ml}^{-1}$ lysate from either *B. wadsworthia*, *Alistipes* or *L. murinus*. Supernatants were analysed for IL-12p40 by ELISA and diluted 25%, then applied to purified CD4⁺ T cells stimulated with plate-bound anti-CD3 ($1 \mu\text{g ml}^{-1}$) and anti-CD28 ($2 \mu\text{g ml}^{-1}$) for 3 days. The supernatants from these cultures were then analysed for IFN- γ and IL-17.

Antibodies and flow cytometry. The following conjugated antibodies were purchased from eBioscience (San Diego): CD4 (GK1.5), CD11c (N418), IFN- γ (XMG1.2), IL-17 (eBio17B7) and isotype controls. Cells were permeabilized with the CytoFix/CytoPerm kit (BD Biosciences) for intra-cytoplasmic detection of IFN- γ and IL-17 cytokines. Flow cytometry analysis was performed with a FACsCanto (BD Biosciences).

26. Rossi, S. S., Converse, J. L. & Hofmann, A. F. High pressure liquid chromatographic analysis of conjugated bile acids in human bile: simultaneous resolution of sulfated and unsulfated lithocholyl amides and the common conjugates bile acids. *J. Lipid Res.* **28**, 589–595 (1987).

SIRT7 links H3K18 deacetylation to maintenance of oncogenic transformation

Matthew F. Barber^{1,2,*†}, Eriko Michishita-Kioi^{2,3,*†}, Yuanxin Xi^{4,*}, Luisa Tasselli^{2,3}, Mitomu Kioi^{5†}, Zarmik Moqtaderi⁶, Ruth I. Tennen^{2,7}, Silvana Paredes^{2,3}, Nicolas L. Young⁸, Kaifu Chen⁴, Kevin Struhl⁶, Benjamin A. Garcia⁸, Or Gozani¹, Wei Li⁴ & Katrin F. Chua^{2,3,7}

Sirtuin proteins regulate diverse cellular pathways that influence genomic stability, metabolism and ageing^{1,2}. SIRT7 is a mammalian sirtuin whose biochemical activity, molecular targets and physiological functions have been unclear. Here we show that SIRT7 is an NAD⁺-dependent H3K18Ac (acetylated lysine 18 of histone H3) deacetylase that stabilizes the transformed state of cancer cells. Genome-wide binding studies reveal that SIRT7 binds to promoters of a specific set of gene targets, where it deacetylates H3K18Ac and promotes transcriptional repression. The spectrum of SIRT7 target genes is defined in part by its interaction with the cancer-associated E26 transformed specific (ETS) transcription factor ELK4, and comprises numerous genes with links to tumour suppression. Notably, selective hypoacetylation of H3K18Ac has been linked to oncogenic transformation, and in patients is associated with aggressive tumour phenotypes and poor prognosis^{3–6}. We find that deacetylation of H3K18Ac by SIRT7 is necessary for maintaining essential features of human cancer cells, including anchorage-independent growth and escape from contact inhibition. Moreover, SIRT7 is necessary for a global hypoacetylation of H3K18Ac associated with cellular transformation by the viral oncoprotein E1A. Finally, SIRT7 depletion markedly reduces the tumorigenicity of human cancer cell xenografts in mice. Together, our work establishes SIRT7 as a highly selective H3K18Ac deacetylase and demonstrates a pivotal role for SIRT7 in chromatin regulation, cellular transformation programs and tumour formation *in vivo*.

The chromatin silencing factor Sir2 (silent information regulator-2) catalyses NAD⁺-dependent histone deacetylation to regulate genomic stability and cellular senescence in budding yeast^{1,2}. In mammals, SIRT7 is the only sirtuin (Sir2 family member) for which a clear enzymatic activity has remained elusive. Indirect evidence has led to the suggestion that SIRT7 deacetylates the tumour suppressor p53 (ref. 7), although *in vitro* and cellular data do not support this model (Supplementary Fig. 1 and ref. 8). In addition, multiple studies have failed to detect direct deacetylase activity of SIRT7 on histones or other substrates¹.

In biochemical fractionation studies, we detected SIRT7 almost exclusively in a chromatin-enriched fraction, suggesting that SIRT7 might function at chromatin (Fig. 1a). We therefore used mass spectrometry to screen for potential NAD⁺-dependent histone deacetylase activity of SIRT7 *in vitro*. Notably, SIRT7 showed highly specific deacetylase activity on peptides containing H3K18Ac, but had no activity on 12 other histone acetylation sites tested (Fig. 1b, c). This striking selectivity of SIRT7 for H3K18Ac contrasts with the broader substrate spectrum of other deacetylases such as SIRT1 (Supplementary

Table 1) or HDAC1 (ref. 9). SIRT7 also showed robust and specific NAD⁺-dependent H3K18Ac-deacetylase activity on full-length histone H3 in purified poly-nucleosomes (Fig. 1d). This activity was abolished by substitution of a conserved histidine residue (H187→Y) in the predicted catalytic domain of SIRT7 and by the general sirtuin inhibitor nicotinamide (Fig. 1e). Selective H3K18Ac deacetylation was also observed in cells after overexpression of the wild-type, but not mutant, SIRT7 protein (Fig. 1f). Moreover, an unbiased proteomic approach using quantitative mass spectrometry independently demonstrated that SIRT7 overexpression induces a dramatic depletion of H3K18Ac in cells (Fig. 1g), whereas changes in other acetylation marks, presumably caused by downstream effects on chromatin structure, were more modest or negligible. Together, our data demonstrate that SIRT7 is an NAD⁺-dependent H3K18Ac deacetylase and the first known deacetylase with high selectivity for the H3K18Ac chromatin mark.

Depletion of H3K18Ac has been associated with aggressive cancer phenotypes and poor patient prognosis^{5,6}, and in cellular studies has been linked to epigenetic reprogramming during transformation of primary human cells by viral oncoproteins^{3,4}. In addition, H3K18Ac is enriched at gene promoters and correlates with transcriptional activation¹⁰. We therefore hypothesized that SIRT7 might deacetylate H3K18Ac at promoters to modulate cancer-related gene expression programs. We first determined the genome-wide occupancy of SIRT7 by chromatin immunoprecipitation (ChIP) sequencing. Results from multiple independent ChIP-sequencing experiments identified 276 SIRT7-binding sites (Supplementary Tables 2–4), which were dramatically enriched for proximal promoter regions (Fig. 2a, b and Supplementary Fig. 2a, b). Notably, SIRT7-binding sites also overlapped significantly with previously mapped regions of H3K18Ac enrichment ($P < 1.4 \times 10^{-80}$) (ref. 10). Together, these data suggest that SIRT7 is a locus-specific enzyme that is positioned to deacetylate H3K18Ac at promoters of a select set of gene targets.

The identified SIRT7 ChIP-sequencing peaks correspond to 241 protein-coding genes (see Methods for details). Using ChIP-quantitative PCR (qPCR) we confirmed the binding of SIRT7 at several of the identified promoters and validated the specificity of the ChIP signals by short interfering RNA (siRNA)-mediated depletion of SIRT7 (Fig. 2c and Supplementary Fig. 3). Functional categorization of the SIRT7-bound genes revealed strong enrichment for factors involved in RNA processing, protein translation and cellular macromolecule metabolism, with diverse links to tumour suppressive activities (Supplementary Fig. 2). Interestingly, SIRT7 bound upstream of several ribosomal protein (RP) genes, whose mis-regulation has

¹Department of Biology, Stanford University, Stanford, California 94305, USA. ²Department of Medicine, Division of Endocrinology, Gerontology, and Metabolism, Stanford University, Stanford, California 94305, USA. ³Geriatric Research, Education, and Clinical Center, VA Palo Alto Health Care System, Palo Alto, California 94304, USA. ⁴Division of Biostatistics, Dan L. Duncan Cancer Center, Department of Molecular and Cellular Biology, Baylor College of Medicine, Houston, Texas 77030, USA. ⁵Department of Radiation Oncology, Stanford University, Stanford, California 94305, USA. ⁶Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts 02115, USA. ⁷Cancer Biology Program, Stanford University, Stanford, California 94305, USA. ⁸Department of Molecular Biology, Princeton University, Princeton, New Jersey 08544, USA. [†]Present addresses: Department of Human Genetics, University of Utah, Salt Lake City, Utah 84112, USA (M.F.B.); R&D Division, Daiichi Sankyo Co., Ltd, Shinagawa-ku, Tokyo 140-8710, Japan (E.M.-K.); Department of Oral and Maxillofacial Surgery, Yokohama City University School of Medicine, 3-9 Fukuura, Kanazawa-ku, Yokohama, 236-0004, Japan (M.K.).

*These authors contributed equally to this work.

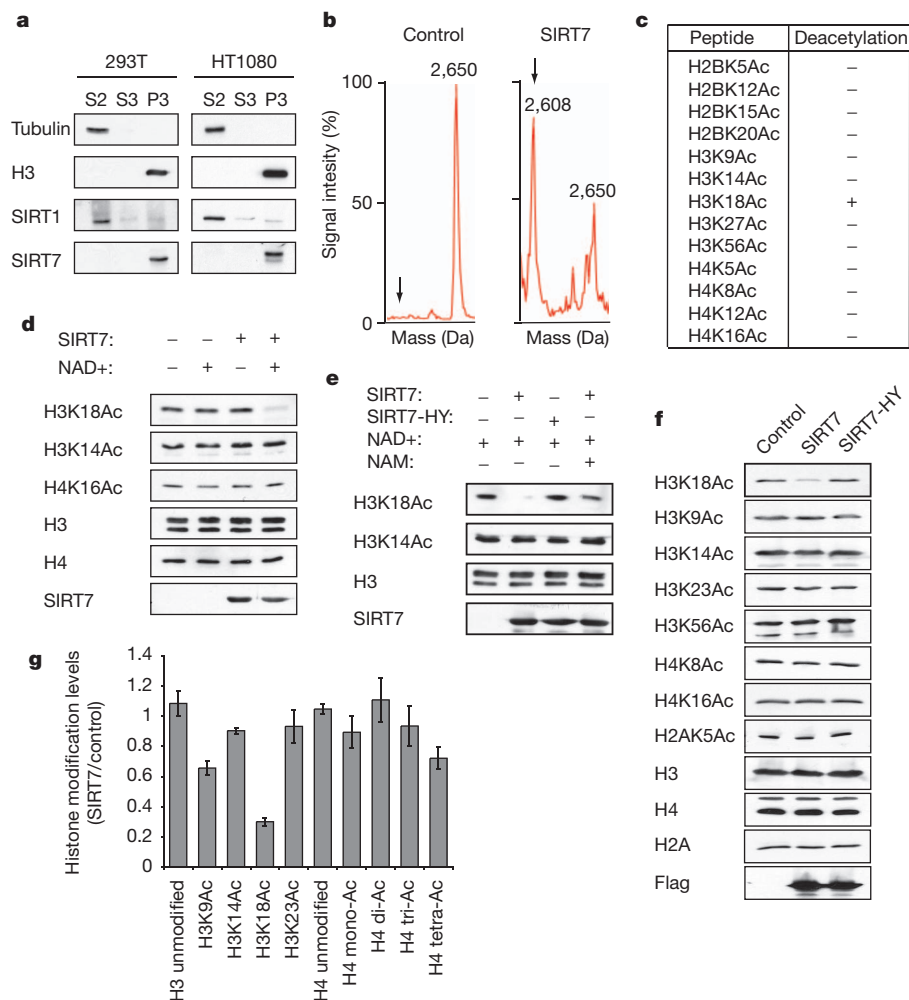


Figure 1 | SIRT7 is a chromatin-associated H3K18Ac-specific deacetylase. **a**, Western blot analysis showing chromatin association of SIRT7 in 293T and HT1080 cells. Biochemical fractions S2, S3 and P3 are enriched for cytoplasm, nucleoplasm or chromatin, respectively. **b**, Mass spectra showing deacetylation of H3K18Ac peptide by SIRT7 compared with negative control reaction lacking enzyme. Molecular masses of acetylated and deacetylated (arrows) peptides are 2,650 and 2,608 daltons, respectively. **c**, Results of SIRT7 deacetylation reactions using acetylated histone peptides, determined by mass spectrometry as in **b**. **d**, **e**, Western blot analysis of H3K18Ac deacetylation activity of wild-type (SIRT7) or mutant (SIRT7-HY) proteins on poly-nucleosomes *in vitro*, and inhibition by nicotinamide (NAM). **f**, Western blot analysis showing H3K18Ac levels in 293T cells transfected with Flag-tagged SIRT7, SIRT7-HY or control empty vector. **g**, Changes in global histone acetylation levels in SIRT7 overexpressing versus control 293T cells, determined by quantitative mass spectrometry. Error bars, s.e.m. of three independent experiments.

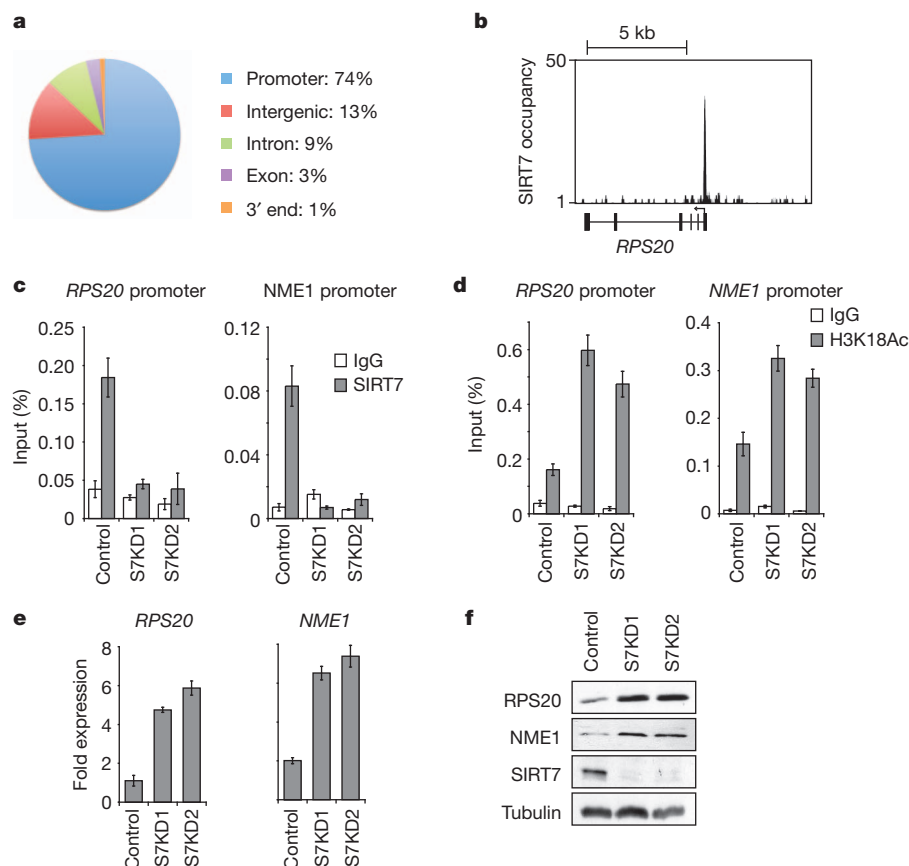


Figure 2 | SIRT7 binds to gene promoters and couples H3K18 deacetylation to transcriptional repression. **a**, Enrichment of SIRT7 in promoter proximal regions, determined by ChIP-seq. **b**, Representative SIRT7 ChIP-seq peak at the *RPS20* gene transcription start site (TSS; arrow). **c**, ChIP-qPCR (mean \pm s.e.m.) showing SIRT7 occupancy in control or SIRT7 knockdown (S7KD1, S7KD2) HT1080 cells, compared with immunoglobulin-G (IgG)-negative control samples. **d**, ChIP-qPCR (mean \pm s.e.m.) showing H3K18Ac hyperacetylation in S7KD HT1080 cells. **e**, Increased expression of SIRT7 target genes in S7KD HT1080 cells determined by qPCR (mean \pm s.e.m.). Signals were normalized to *GAPDH* expression. **f**, Western blots of cell extracts corresponding to samples in **e**.

been linked to cancer in multiple settings (see below), as well as genes found repressed in aggressive cancers or identified in screens for tumour suppressor genes (for example, *NME1* and *COPS2*)^{11,12}.

We next asked whether SIRT7 deacetylates H3K18Ac at the promoters of specific candidate target genes. SIRT7 depletion led to hyperacetylation of H3K18 at the promoters of the *RPS20*, *RPS7*, *RPS14*, *NME1* and *COPS2* genes, but not multiple negative control promoters (Fig. 2d and Supplementary Fig. 4). Consistent with this locus specificity, global H3K18Ac levels were not affected by SIRT7 depletion (Supplementary Fig. 5). Importantly, SIRT7 knockdown (S7KD) also led to specific increases in expression of multiple target genes (Fig. 2e, f and Supplementary Figs 6 and 7), whereas depletion of HDAC1 (which can also deacetylate H3K18Ac⁹) did not (Supplementary Fig. 8). Together, our findings demonstrate that SIRT7 functions in gene-specific transcriptional repression at a select subset of H3K18Ac-containing promoters.

We next asked how selective recruitment of SIRT7 to its target promoters is achieved. SIRT7 lacks known sequence-specific DNA-binding domains, leading us to hypothesize that it might interact with other proteins that contain such domains. We therefore identified *de novo* DNA motifs that are enriched in SIRT7-bound promoter sequences, and compared these motifs with curated transcription-factor binding motifs in the JASPAR CORE database¹³. Of the 50 most significant SIRT7-associated motifs, 25 corresponded to consensus-binding sites for the ETS family of transcription factors, many of which are important modulators of cellular transformation and cancer progression¹⁴. The SIRT7 consensus motif was most similar to the DNA sequence recognized by the ETS protein ELK4 (Fig. 3a).

Although its molecular function has not been extensively studied, ELK4 has been implicated in both transcriptional activation and repression¹⁵. Of the 276 SIRT7-binding sites that we identified by ChIP-sequencing, 57.6% contain at least one ELK4 consensus motif, a significant enrichment over total RefSeq promoters ($P < 3.1 \times 10^{-89}$) (Supplementary Tables 5, 6). In addition, approximately 70% of SIRT7-binding sites overlap with ELK4 peaks previously identified by ChIP-sequencing ($P < 1 \times 10^{-300}$) (ref. 16). To examine the potential interplay between SIRT7 and ELK4, we first confirmed that ELK4 binds several of the SIRT7 target promoters that contain the ETS consensus motif (*NME1* and *COPS2*), but not promoters lacking the motif (*RPS20* or *GAPDH*; Supplementary Fig. 9a). Next, in co-immunoprecipitation experiments, we found that SIRT7 interacts physically with ELK4 (Fig. 3b, c and Supplementary Fig. 10a), but not with two other ETS proteins, ELK1 and GABP- α (Supplementary Fig. 10b). To assess the functional importance of this interaction, we examined the effects of ELK4 knockdown on SIRT7 ChIP occupancy at specific target promoters. Depletion of ELK4 led to a partial but significant decrease in SIRT7 occupancy at the *NME1* and *COPS2* promoters but not the *RPS20* promoter (Fig. 3d, e and Supplementary Fig. 11a, b), and did not alter global levels of SIRT7 at chromatin (Supplementary Fig. 12). Moreover, ELK4 knockdown led to elevated H3K18Ac levels at the *NME1* and *COPS2* promoters, but not at promoters lacking the ETS motif (Supplementary Fig. 9b). Together, these findings suggest that ELK4 functions to target SIRT7 to specific promoters for H3K18 deacetylation.

We next examined the effects of ELK4 knockdown on gene repression by SIRT7. ELK4 knockdown did not appreciably alter expression of *NME1* and *COPS2* under baseline conditions (Fig. 3f), probably because considerable SIRT7 protein remained bound at these promoters (Fig. 3e). This incomplete loss of promoter-bound SIRT7 could reflect the incomplete depletion of ELK4 (Fig. 3d) as well as compensatory activity of other ETS factors in SIRT7 targeting. Indeed, the SIRT7 ChIP-sequencing peaks displayed some overlap (25%; $P < 1 \times 10^{-300}$) with binding sites for ELK1. Importantly, however, knockdown of ELK4, but not ELK1 or GABP- α , was sufficient to impair gene repression induced by overexpression of SIRT7 (Fig. 3f and Supplementary Figs 11c and 13). Thus, any compensatory capacity of other ETS factors

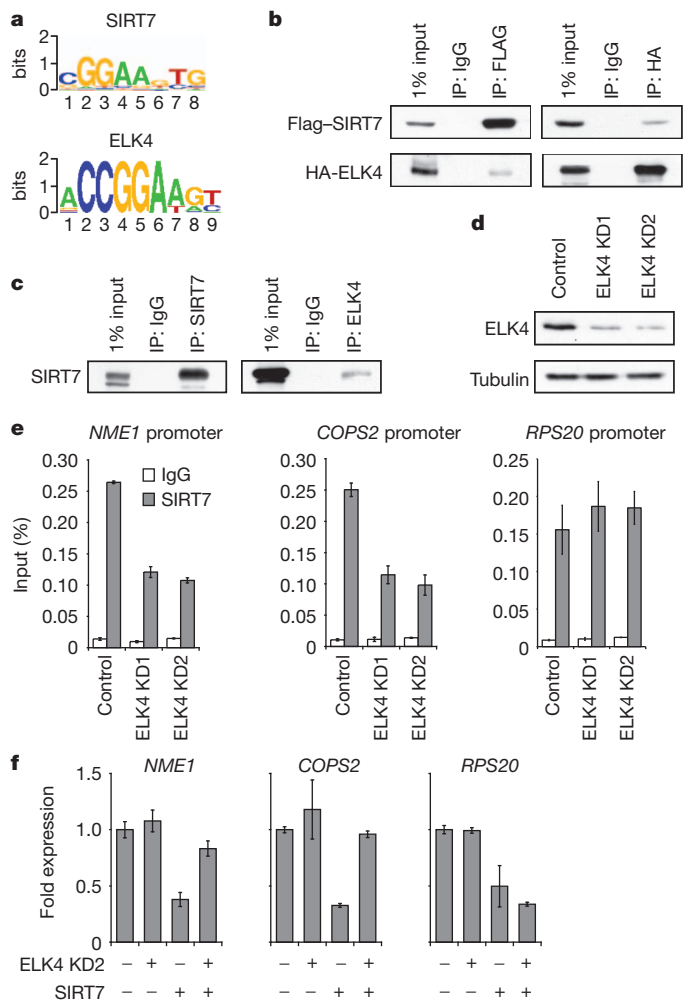


Figure 3 | SIRT7 is stabilized at target promoters by interaction with the ETS family transcription factor ELK4. **a**, Comparison of the SIRT7 consensus motif to the ELK4 consensus motif ($P < 9.66 \times 10^{-9}$). **b**, Western blot analysis showing co-immunoprecipitation of Flag-tagged SIRT7 and HA-tagged ELK4 expressed in 293T cells. **c**, Western blots showing co-immunoprecipitation (IP) of endogenous SIRT7 and ELK4 proteins. **d**, Western blots showing knockdown of ELK4 from HT1080 cells with two independent siRNAs. **e**, Partial reduction of SIRT7 occupancy at target promoters in ELK4 KD HT1080 cells determined by ChIP (mean \pm s.e.m.). **f**, ELK4 depletion attenuates SIRT7-mediated transcriptional repression in HT1080 cells, as determined by qPCR. Error bars, s.e.m. of three independent experiments.

is exceeded under conditions of elevated SIRT7 expression, and in this setting ELK4 is the main ETS factor responsible for SIRT7 targeting. Thus, we conclude that the promoter stabilization of SIRT7 by ELK4 is important for SIRT7-mediated gene repression. Moreover, this functional interplay between ELK4 and SIRT7 might be particularly important in settings of elevated SIRT7 expression, which occurs in certain cancers (Supplementary Fig. 14)^{17–19}.

Analysis of SIRT7-occupied genes revealed a clear correlation with factors whose expression is altered in various cancers (Supplementary Fig. 2d). This observation, together with previous reports linking both H3K18Ac^{3–6} and ELK4 (ref. 20) to cancer, suggested that SIRT7 might regulate transformed features of cancer cells. Indeed, SIRT7 depletion in HT1080 and U2OS cells severely impaired both anchorage-independent cellular growth in soft agar and proliferation in low serum, two important hallmarks of transformed cells (Fig. 4a, b and Supplementary Figs 15 and 16). These effects of SIRT7-depletion were associated with both increased cell death and altered cell-cycle progression (Supplementary Fig. 17), and were observed in prostate cancer cells, a setting where overexpression of both ELK4 and SIRT7

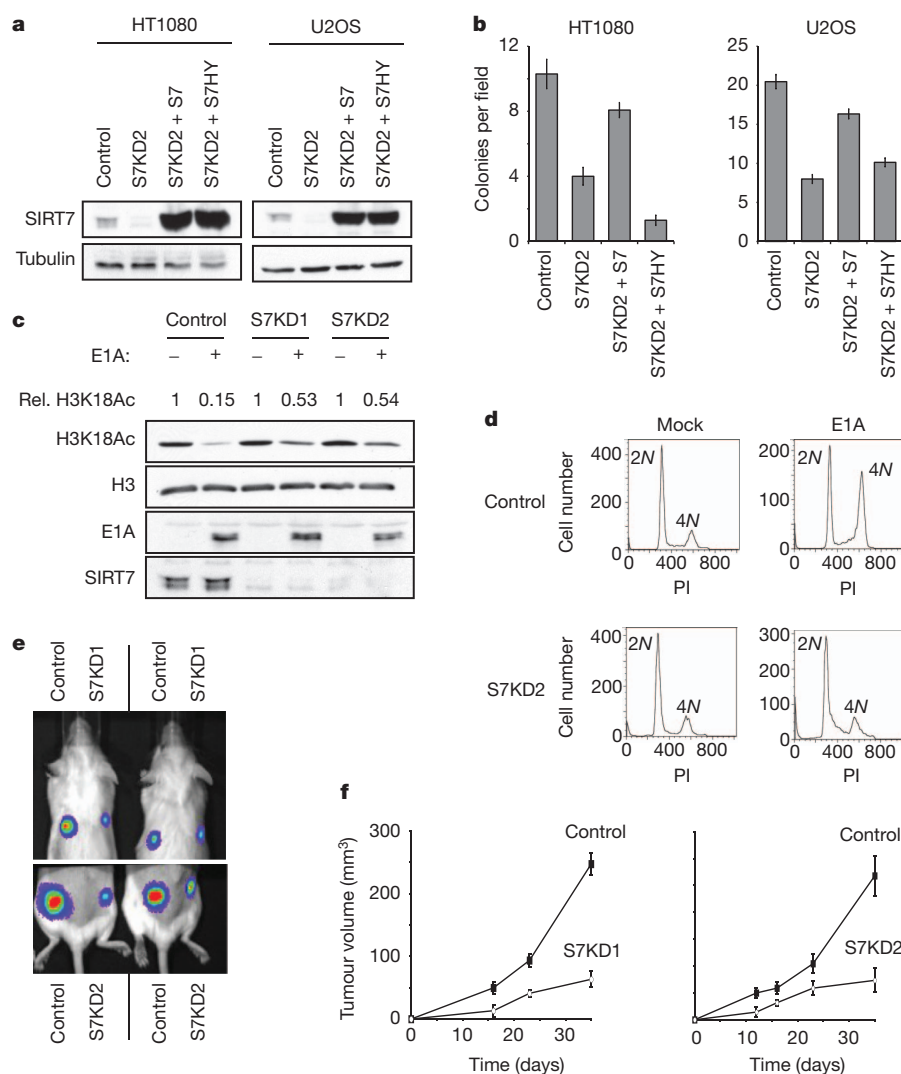


Figure 4 | SIRT7 depletion reverses cancer cell phenotypes and inhibits tumour growth *in vivo*. **a**, Western blots showing SIRT7 levels from stable cell lines used in **b**. **b**, Reduced anchorage-independent growth of SIRT7 knockdown cells when plated in soft agar, and reconstitution with wild-type but not mutant SIRT7. Data represent mean \pm s.e.m. of three independent experiments. **c**, Western blot analysis showing impaired H3K18Ac deacetylation induced by E1A in SIRT7 knockdown HT1080 cells. Rel. H3K18Ac, relative levels of H3K18Ac in mock-treated versus E1A expressing cells, normalized to total H3 levels. **d**, SIRT7 depletion impairs E1A-mediated loss of contact inhibition in primary IMR90 fibroblasts determined by flow cytometry. DNA content (2N or 4N, as determined by propidium iodide (PI) staining) is indicated. **e**, Representative imaging of tumours derived from SIRT7 knockdown or control cells, following subcutaneous xenograft transplants in immunodeficient mice, 16 days after injection. **f**, Tumour volume (mean \pm s.e.m.; $n = 5$) as in **e**, measured over 35 days.

has been observed (Supplementary Fig. 14)²⁰. Importantly, functional reconstitution assays revealed that the catalytic activity of SIRT7 is necessary for maintenance of the cancer cell-specific growth properties (Fig. 4a, b and Supplementary Fig. 16), linking the biochemical activity and cancer-related functions of SIRT7. In addition, simultaneous expression of SIRT7 and ELK4 had a synergistic effect on maintenance of the transformed phenotype of these cells (Supplementary Fig. 18), further highlighting the importance of the molecular interplay between SIRT7 and ELK4.

The adenoviral E1A oncoprotein induces a specific decrease in H3K18 acetylation that is important for its transforming activity^{3,4}. Strikingly, SIRT7 depletion in HT1080 cells severely inhibited this E1A-dependent reduction of H3K18Ac (Fig. 4c). Moreover, expression of E1A in non-dividing, contact-inhibited primary human fibroblasts triggers cell-cycle re-entry and escape from contact inhibition, another hallmark of oncogenic transformation²¹, and SIRT7 depletion abolished this effect (Fig. 4d). Thus, SIRT7 is required for both the global H3K18Ac deacetylation and escape from contact inhibition that are induced by the E1A oncoprotein. Finally, we examined the effect of SIRT7-knockdown on tumour growth using subcutaneous xenografts of U251 cancer cells in mice, and found that tumour formation was severely impaired by SIRT7 depletion (Fig. 4e, f and Supplementary Fig. 19). Together, our data suggest that H3K18Ac-specific deacetylation by SIRT7 is important for maintaining fundamental properties of the cancer cell phenotype and stabilizing the tumorigenicity of human cancer cells *in vivo*.

In summary, we have demonstrated that SIRT7 is a promoter-associated, highly selective H3K18Ac deacetylase that mediates transcriptional repression and stabilizes cancer cell phenotypes. These findings suggest that pathological upregulation of SIRT7 in cancer cells may contribute to the malignant phenotype of certain tumours. Indeed, SIRT7 overexpression is observed in multiple cancer tissues (Supplementary Fig. 14)^{17–19}, and the cBio Cancer Genomics Portal has reported 55 separate instances of SIRT7 gene amplification in patients' tumours so far (<http://www.cbioportal.org>). We note that although SIRT7 is important for maintaining the transformed state of cancer cells, we have not observed a role for SIRT7 in initiating the process of cellular transformation itself. For example, overexpression of SIRT7 in immortalized mouse embryonic fibroblasts or primary human fibroblasts did not lead to oncogenic transformation (data not shown, and ref. 8). Thus, our data suggest models in which H3K18Ac deacetylation by SIRT7 modulates the epigenetic stability and tumorigenicity of cancer cells, but how SIRT7 deficiency impacts on tumour initiation and the overall incidence of cancer is probably more complex.

Our observation that SIRT7 represses several ribosomal protein genes is intriguing, as mutations in ribosomal protein genes have been linked to cancer progression^{22,23}. For example, the SIRT7 target gene *RPS14* is a disease gene of the human 5q[–] syndrome, a myelodysplastic disorder that frequently progresses to acute myeloid leukaemia²³, and multiple ribosomal proteins have been identified as haploinsufficient tumour suppressors in zebrafish²². The molecular mechanisms underlying the links between ribosomal protein insufficiency and cancer are

unclear, but have been hypothesized to involve imbalances in translation regulation or translation-independent functions of individual ribosomal proteins^{22,23}.

Previous studies have found that SIRT7 promotes ribosomal RNA transcription²⁴, although this function appears to be specific to cell type or experimental conditions (Supplementary Fig. 20). Whether such an activity functions in parallel to, or as a consequence of, SIRT7's role in ribosomal protein gene repression remains to be elucidated, but may suggest a broad role for SIRT7 in coordinating the cellular translation machinery. Interestingly, ribosomal protein gene deletions and inhibition of translation have also been linked to lifespan extension in numerous model organisms, including mammals^{25,26}, suggesting that gene repression by SIRT7 might also influence ageing-related cellular processes. Consistent with this hypothesis, one strain of *Sirt7*-deficient mice shows cardiac defects and shortened lifespan⁸, although this phenotype appears to depend on genetic background²⁷. Future work should shed light on the potential role of SIRT7 in ageing-associated pathologies and lifespan determination.

METHODS SUMMARY

Histone deacetylation assays. *In vitro* histone deacetylation assays were performed as previously described²⁸ using acetylated peptides or poly-nucleosomes purified from HeLa cells as substrate. Recombinant human SIRT7 protein was purified from baculovirus-infected insect cells as described⁸.

ChIP-qPCR and messenger RNA analysis. ChIP was performed as previously described²⁹, except that the Qiagen PCR purification kit was used for DNA purification (Qiagen). Whole messenger RNA was purified from cells using the RNeasy Mini Kit (Qiagen). Quantitative reverse transcription PCR was performed using the Universal ProbeLibrary System with a LightCycler 480 II (Roche), or using Taqman Gene Expression Assays on a 7300 Real Time PCR machine (Applied Biosystems). RNA from patient-matched tumour and unaffected control tissues was purchased from Ambion.

Tumour xenograft experiments. Equal numbers of U251 cells expressing luciferase and either control (pSR) or SIRT7 knockdown vectors (upper quadrants, 4×10^6 pSR or S7KD1 cells; lower quadrants, 8×10^6 pSR or S7KD2 cells) were implanted on the backs of RAG knockout mice. Tumour growth was monitored using callipers and visualized with a bioluminescence-based IVIS system (Caliper LifeSciences).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 25 March 2011; accepted 14 March 2012.

Published online 6 May 2012.

- Haigis, M. C. & Sinclair, D. A. Mammalian sirtuins: biological insights and disease relevance. *Annu. Rev. Pathol.* **5**, 253–295 (2010).
- Longo, V. D. & Kennedy, B. K. Sirtuins in aging and age-related disease. *Cell* **126**, 257–268 (2006).
- Ferrari, R. *et al.* Epigenetic reprogramming by adenovirus e1a. *Science* **321**, 1086–1088 (2008).
- Horwitz, G. A. *et al.* Adenovirus small e1a alters global patterns of histone modification. *Science* **321**, 1084–1085 (2008).
- Manuyakorn, A. *et al.* Cellular histone modification patterns predict prognosis and treatment response in resectable pancreatic adenocarcinoma: results from RTOG 9704. *J. Clin. Oncol.* **28**, 1358–1365 (2010).
- Seligson, D. B. *et al.* Global levels of histone modifications predict prognosis in different cancers. *Am. J. Pathol.* **174**, 1619–1628 (2009).
- Vakhrusheva, O. *et al.* Sirt7 increases stress resistance of cardiomyocytes and prevents apoptosis and inflammatory cardiomyopathy in mice. *Circ. Res.* **102**, 703–710 (2008).
- Michishita, E., Park, J. Y., Burnes, J. M., Barrett, J. C. & Horikawa, I. Evolutionarily conserved and nonconserved cellular localizations and functions of human SIRT proteins. *Mol. Biol. Cell* **16**, 4623–4635 (2005).
- Hassig, C. A. *et al.* A role for histone deacetylase activity in HDAC1-mediated transcriptional repression. *Proc. Natl Acad. Sci. USA* **95**, 3519–3524 (1998).
- Wang, Z. *et al.* Combinatorial patterns of histone acetylations and methylations in the human genome. *Nature Genet.* **40**, 897–903 (2008).
- Steeg, P. S. *et al.* Evidence for a novel gene associated with low tumor metastatic potential. *J. Natl. Cancer Inst.* **80**, 200–204 (1988).

- Leal, J. F. *et al.* Cellular senescence bypass screen identifies new putative tumor suppressor genes. *Oncogene* **27**, 1961–1970 (2008).
- Bryne, J. C. *et al.* JASPAR, the open access database of transcription factor-binding profiles: new content and tools in the 2008 update. *Nucleic Acids Res.* **36**, D102–D106 (2008).
- Galang, C. K., Muller, W. J., Foos, G., Oshima, R. G. & Hauser, C. A. Changes in the expression of many Ets family transcription factors and of potential target genes in normal mammary tissue and tumors. *J. Biol. Chem.* **279**, 11281–11292 (2004).
- Kaikkonen, S., Makkonen, H., Rytinki, M. & Palvimäki, J. J. SUMOylation can regulate the activity of ETS-like transcription factor 4. *Biochim. Biophys. Acta* **1799**, 555–560 (2010).
- O'Geen, H. *et al.* Genome-wide binding of the orphan nuclear receptor TR4 suggests its general role in fundamental biological processes. *BMC Genomics* **11**, 689 (2010).
- Ashraf, N. *et al.* Altered sirtuin expression is associated with node-positive breast cancer. *Br. J. Cancer* **95**, 1056–1061 (2006).
- de Nigris, F. *et al.* Isolation of a SIR-like gene, SIR-T8, that is overexpressed in thyroid carcinoma cell lines and tissues. *Br. J. Cancer* **86**, 917–923 (2002).
- Frye, R. 'SIRT8' expressed in thyroid cancer is actually SIRT7. *Br. J. Cancer* **87**, 1479 (2002).
- Makkonen, H. *et al.* Identification of ETS-like transcription factor 4 as a novel androgen receptor target in prostate cancer cells. *Oncogene* **27**, 4865–4876 (2008).
- Braithwaite, A. W. *et al.* Adenovirus-induced alterations of the cell growth cycle: a requirement for expression of E1A but not of E1B. *J. Virol.* **45**, 192–199 (1983).
- Amsterdam, A. *et al.* Many ribosomal protein genes are cancer genes in zebrafish. *PLoS Biol.* **2**, E139 (2004).
- Ebert, B. L. *et al.* Identification of RPS14 as a 5q- syndrome gene by RNA interference screen. *Nature* **451**, 335–339 (2008).
- Ford, E. *et al.* Mammalian Sir2 homolog SIRT7 is an activator of RNA polymerase I transcription. *Genes Dev.* **20**, 1075–1080 (2006).
- Hansen, M. *et al.* Lifespan extension by conditions that inhibit translation in *Caenorhabditis elegans*. *Aging Cell* **6**, 95–110 (2007).
- Harrison, D. E. *et al.* Rapamycin fed late in life extends lifespan in genetically heterogeneous mice. *Nature* **460**, 392–395 (2009).
- Lombard, D. B., Schwer, B., Alt, F. W. & Mostoslavsky, R. SIRT6 in DNA repair, metabolism and ageing. *J. Intern. Med.* **263**, 128–141 (2008).
- Michishita, E. *et al.* SIRT6 is a histone H3 lysine 9 deacetylase that modulates telomeric chromatin. *Nature* **452**, 492–496 (2008).
- Dahl, J. A. & Collas, P. Q2ChIP, a quick and quantitative chromatin immunoprecipitation assay, unravels epigenetic dynamics of developmentally regulated genes in human carcinoma cells. *Stem Cells* **25**, 1037–1046 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank M. Snyder and colleagues for high-throughput sequencing (conducted as part of the ENCODE consortium), and members of the Chua and Gozani laboratories for discussions and comments on the manuscript. This work was supported by grants from the National Institutes of Health (NIH) to K.F.C. (K08 AG028961, R01 AG028867), W.L. (U01DA025956), O.G. (R01 GM079641), K.S. (GM 30186, HG 4558) and B.A.G. (DP2OD007447); from the National Science Foundation to B.A.G. (CAREER Award, CBET-0941143); from the Department of Defense to W.L. (PC094421); from the Cancer Prevention and Research Institute of Texas (CPRIT) to W.L. (RP110471-C3); from the Department of Veterans Affairs to K.F.C. (Merit Award); and by fellowship awards to M.F.B. (ARCS Scholarship and Mason Case Graduate Fellowship), R.I.T. (NIH training grant 1018438-142 PABCA), L.T. (American Italian Cancer Foundation Post-doctoral Research Fellowship), S.P. (NIH training grant 3T32DK007217-36S1) and N.L.Y. (NIH F32 NRSA). W.L. is a recipient of a Duncan Scholar Award. K.F.C. is a Paul Beeson Scholar and an Ellison Medical Foundation New Scholar in Aging.

Author Contributions M.F.B. and K.F.C. conceived the project and, with E.M.-K. and O.G., designed the experiments. M.F.B. and E.M.-K. performed and interpreted molecular and cell biology experiments, and M.F.B. and K.F.C. wrote the manuscript with input from co-authors. Y.X., K.C. and W.L. performed the bioinformatic analyses and wrote the corresponding manuscript sections. Z.M. and K.S. designed and performed the SIRT7 ChIP-sequencing experiments. L.T. performed and interpreted the experiments in Supplementary Fig. 17; B.A.G. and N.L.Y. performed the quantitative mass spectrometry in Fig. 1g; M.K. performed the mouse xenograft experiments in Fig. 4; R.I.T. and S.P. generated constructs for various experiments and provided technical assistance. M.F.B., E.M.-K. and Y.X. made independent contributions to the work.

Author Information The SIRT7 ChIP-sequencing data are deposited in NIH Gene Expression Omnibus under accession number GSE28149. In addition, raw and processed data are available on our project website, <http://dlccc-web.brc.bcm.edu/lilab/SIRT7/>. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to K.F.C. (kfchua@stanford.edu) or W.L. (WL1@bcm.edu).

METHODS

Cell culture, RNA interference and viral transduction. Human 293T, HT1080, U251 and U2OS cell lines were acquired from the American Type Culture Collection (ATCC), and DU145 cells were a gift from P. Khavari. These cells were cultured in Advanced DMEM (Invitrogen) supplemented with penicillin–streptomycin (Invitrogen), GlutaMAX-1 (Invitrogen) and 10% newborn calf serum. K562 cells were cultured in RPMI (Invitrogen) supplemented with penicillin–streptomycin (Invitrogen), GlutaMAX-1 (Invitrogen) and 10% newborn calf serum. IMR90 cells were cultured in DMEM/F12 (Invitrogen) containing penicillin–streptomycin, GlutaMAX-1 and 10% fetal bovine serum. Retroviral transduction was performed as previously described²⁸. SIRT7 knockdown target sequences were as follows: S7KD1, 5'-CACCTTTCTGTGAGAAGGAA-3'; S7KD2, 5'-TAGCCATTTGCTCTGAGGAA-3'; S7KD3, 5'-GCCTGAAGGTTCTAAAGAA-3'; S7KD4, 5'-GAACGGAAGCTCGGGTTATT-3'. ELK4 knockdown target sequences were as follows: ELK4 KD1, 5'-CGACACAGACATTGATTCA-3'; ELK4 KD2, 5'-GAGAATGGAGGGAAAGATA-3', as previously described²⁰. ELK1 knockdown target sequence was 5'-GATGTGAGTAGAAGAGTTA-3'. GABP α knockdown target sequence was 5'-TGAAGAAGCTCAAGTGATA-3'. HDAC1 knockdown target sequences were HDAC1 KD1 5'-AGAAAGACCCAGAGGAGAA-3', HDAC1 KD2 5'-GCAAGCAGATGCAGAGATT-3'. Double-stranded siRNAs were purchased from Thermo Scientific. For retroviral packaging, 293T cells were co-transfected with pVPack-VSV-G, pVPack-GP (Stratagene) and the SIRT7 knockdown or pSUPERretro control constructs, and viral supernatant was collected after 48 h. For transduction, cells were incubated with virus-containing supernatant in the presence of $8 \mu\text{g ml}^{-1}$ polybrene. After 48 h, infected cells were selected for 72 h with puromycin ($2 \mu\text{g ml}^{-1}$) or hygromycin ($200 \mu\text{g ml}^{-1}$). Antibodies and PCR primer details are provided in Supplementary Tables 7 and 8. Adenovirus expressing the small E1A gene alone was generated and used to infect IMR90 cells using the Virapower Adenovirus System (Invitrogen) according to the manufacturer's instructions. Anchorage-independent growth was measured as previously described³⁰. Annexin V analysis was performed using the FITC Annexin Apoptosis Detection Kit (BD Pharmingen), on S7KD and control U2OS cells cultured in 1% serum. For cell cycle analysis, S7KD and control HT1080 cells were pulsed with $33 \mu\text{M}$ BrdU, fixed in 75% ethanol in PBS, stained with fluorescein isothiocyanate (FITC) mouse anti-BrdU (BD Pharmingen) and propidium iodide, as previously described²⁸. Flow cytometry data were acquired using a fluorescence-activated cell sorting (FACS) LSRFortessa flow cytometer and FACS Diva software (BD Biosciences), and analysed with CellQuest-Pro software (BD Biosciences). For analysis of H3K18Ac in E1A expressing cells, HT1080 cells were treated with control or SIRT7 siRNAs for 24 h, then transfected with control (empty vector) or E1A-expressing vectors. Forty-eight hours after siRNA transfection, extracts were prepared and analysed by western blot. Relative levels of H3K18Ac (rel. H3K18Ac) were determined by quantifying H3K18Ac western blot band intensities using ImageJ software, and normalizing to total H3 band intensities. Samples expressing E1A were set relative to their matched control.

Biochemical fractionation and co-immunoprecipitations. Samples enriched for cytoplasmic, nucleoplasmic and chromatin fractions were prepared as previously described³¹. Co-immunoprecipitations were performed as previously described³², except that one 150 mm plate of cells was used per immunoprecipitation. Protein A/G beads (Sigma) were used instead of Flag-resin, and elution was performed by boiling beads in Laemmli loading buffer.

Histone deacetylation assays. *In vitro* histone deacetylation assays were performed as previously described²⁸. Purification of human SIRT7 protein from baculovirus-infected insect cells was described previously⁸. Calf thymus histones were obtained from Roche, and poly-nucleosomes were purified from HeLa cells as previously described³³. Histone peptides were synthesized at the Yale W. M. Keck peptide synthesis facility, and liquid chromatography mass spectrometry was performed at the Stanford University Vincent Coates Foundation Mass Spectrometry Laboratory. To determine histone acetylation levels in cells, 293T cells were transiently transfected with pcDNA 3.1 vectors containing Flag-tagged wild-type SIRT7, the SIRT7-HY catalytic mutant or an empty vector. Whole-cell lysates were collected after 48 h. Western blot analysis of histone acetylation levels was performed with modification-specific antibodies.

Quantitative mass spectrometry. Acid-extracted total histones were subjected to chemical derivatization using D₀-propionic anhydride and digested with trypsin at a substrate to enzyme ratio of 10:1 for 6 h at 37 °C as previously described³⁴. An additional round of propionylation was performed on the digested peptides, with one sample being derivatized with the same D₀-propionic anhydride reagent, and the other being derivatized with D₁₀-propionic anhydride for quantitative proteomics as previously described³⁵. D₁₀-propionic anhydride introduces a 5 dalton shift by derivatization of the free amino termini of all peptides generated from the trypsin digest. Equal amounts of both samples as quantified earlier by a

Bradford assay were mixed together, and digested peptides were de-salted using homemade STAGE tips as reported earlier³⁶. Desalted peptides were loaded onto fused silica microcapillary column (75 μm) packed with C18 resin constructed with an ESI tip through an Eksigent AS-2 autosampler (Eksigent Technologies) at a rate of approximately 200 nl min^{-1} . Peptides were eluted using a 5–35% solvent B gradient for 60 min (solvent A = 0.1 M acetic acid, solvent B = 70% MeCN in 0.1 M acetic acid). Nanoflow liquid chromatography coupled with tandem mass spectrometry (LC–MS/MS) experiments were performed on an Orbitrap mass spectrometer (ThermoFisher Scientific) taking a full mass spectrum at 30,000 resolution in the Orbitrap and seven data-dependent MS/MS spectra in the ion trap. All MS and MS/MS spectra were manually verified and quantified.

ChIP-seq and computational analysis. ChIP for ChIP-seq analysis was performed as previously described³⁷. Four ChIP samples were sequenced using Illumina Solexa Genome Analyzer II single-end sequencing protocol, including two SIRT7 replicates and two input control replicates. Sequencing adapters and low-quality reads were removed, and the trimmed reads were aligned to human reference genome hg18 by GATK data processing pipeline, allowing up to two mismatches. The biological replicates of SIRT7 ChIP-seq were first analysed individually to measure the reproducibility. The result indicated that the two biological replicates were very similar and met all the NIH ENCODE data quality guidelines for high reproducibility (Supplementary Table 4). The uniquely mapped reads from replicates of SIRT7 and input control samples were pooled respectively and processed by MACS (version 1.3.6)³⁸ to generate the whole-genome ChIP-seq profiles, with the '-diag' option enabled for the sequencing depth saturation test. Clonal reads were automatically removed by MACS. In total, 276 SIRT7-binding sites were identified with *P* value cut off 1×10^{-8} . The wig files were normalized to 10,000,000 total tag number and converted into bigwig format for visualization. The SIRT7 target genes were identified by detecting the SIRT7-binding peaks within 3 kilobases (kb) upstream to 3 kb downstream of TSSs of RefSeq genes using CEAS³⁵. In total, 253 target genes were identified, including 241 protein-coding genes. The gene ontology analysis was performed using DAVID Bioinformatics Resources 6.7 (<http://david.abcc.ncifcrf.gov>)^{39,40}.

Cancer gene association was performed using the Oncomine database (<http://www.oncomine.org>). For identification of ELK4 ChIP-seq target genes, the ChIP-seq reads from O'Geen *et al.* (GSE24685)¹⁶ were remapped to hg18, and peaks were called using MACS (Supplementary Table 9). The target genes were identified by searching for ELK4 peaks 3 kb up- and downstream of TSSs.

De novo motifs with sizes from 6 to 15 nucleotides were searched within SIRT7-binding sites using MDModule⁴¹, with repetitive regions masked and running parameters '-s 100 -t 50'. The top 50 detected *de novo* motifs (top five of each motif size) were recorded and compared with JASPAR motif database using STAMP with default settings⁴². The position weight matrix of the ELK4 motif (Supplementary Table 6) was remapped to the identified SIRT7 peaks using cisgenome³⁴, with parameter '-r 30'.

ChIP and mRNA analysis. Cells were prepared for ChIP as previously described²⁹, with the exception that DNA was washed and eluted using a PCR purification kit (Qiagen) rather than by phenol–chloroform extraction. Whole mRNA was purified from cells using the RNeasy Mini Kit (Qiagen). Quantitative reverse transcription PCR was performed using the Roche Universal ProbeLibrary System with a LightCycler 480 II (Roche), or using Taqman Gene Expression Assays (Applied Biosystems) on a 7300 Real Time PCR machine (Applied Biosystems). Pre-rRNA custom primer–probe mix was generated by Applied Biosystems using human pre-rRNA DNA sequence. RNA from patient-matched tumour and unaffected control tissues was purchased from Ambion.

Tumour xenograft experiments. Equal numbers of U251 cells expressing luciferase and either control (pSR) or SIRT7 knockdown (S7KD) vectors (upper quadrants: 4×10^6 pSR or S7KD1 cells; lower quadrants: 8×10^6 pSR or S7KD2 cells) were implanted on the backs of RAG knockout mice. Tumour growth was monitored using callipers and visualized with a bioluminescence-based IVIS system (Caliper LifeSciences).

30. Chua, K. F. *et al.* Mammalian SIRT1 limits replicative life span in response to chronic genotoxic stress. *Cell Metab.* **2**, 67–76 (2005).
31. Mendez, J. & Stillman, B. Chromatin association of human origin recognition complex, cdc6, and minichromosome maintenance proteins during the cell cycle: assembly of prereplication complexes in late mitosis. *Mol. Cell. Biol.* **20**, 8602–8612 (2000).
32. Hung, T. *et al.* ING4 mediates crosstalk between histone H3 K4 trimethylation and H3 acetylation to attenuate cellular transformation. *Mol. Cell* **33**, 248–256 (2009).
33. Shi, X. *et al.* ING2 PHD domain links histone H3 lysine 4 methylation to active gene repression. *Nature* **442**, 96–99 (2006).
34. Garcia, B. A. *et al.* Chemical derivatization of histones for facilitated analysis by mass spectrometry. *Nature Protocols* **2**, 933–938 (2007).

35. Plazas-Mayorca, M. D. *et al.* One-pot shotgun quantitative mass spectrometry characterization of histones. *J. Proteome Res.* **8**, 5367–5374 (2009).
36. Rappsilber, J., Ishihama, Y. & Mann, M. Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Anal. Chem.* **75**, 663–670 (2003).
37. Moqtaderi, Z. *et al.* Genomic binding profiles of functionally distinct RNA polymerase III transcription complexes in human cells. *Nature Struct. Mol. Biol.* **17**, 635–640 (2010).
38. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
39. Dennis, G. Jr *et al.* DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol.* **4**, 3 (2003).
40. Huang da, W. Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols* **4**, 44–57 (2009).
41. Liu, X. S., Brutlag, D. L. & Liu, J. S. An algorithm for finding protein-DNA binding sites with applications to chromatin-immunoprecipitation microarray experiments. *Nature Biotechnol.* **20**, 835–839 (2002).
42. Mahony, S. & Benos, P. V. STAMP: a web tool for exploring DNA-binding motif similarities. *Nucleic Acids Res.* **35**, W253–W258 (2007).

SbsB structure and lattice reconstruction unveil Ca^{2+} triggered S-layer assembly

Ekaterina Baranova^{1,2}, Rémi Fronzes^{3,4}, Abel Garcia-Pino², Nani Van Gerven^{1,2}, David Papapostolou⁵, Gérard Péhau-Arnaudet⁴, Els Pardon², Jan Steyaert², Stefan Howorka⁵ & Han Remaut^{1,2}

S-layers are regular two-dimensional semipermeable protein layers that constitute a major cell-wall component in archaea and many bacteria^{1–3}. The nanoscale repeat structure of the S-layer lattices and their self-assembly from S-layer proteins (SLPs) have sparked interest in their use as patterning and display scaffolds for a range of nano-biotechnological applications^{4–7}. Despite their biological abundance and the technological interest in them, structural information about SLPs is limited to truncated and assembly-negative proteins^{8–10}. Here we report the X-ray structure of the SbsB SLP of *Geobacillus stearothermophilus* PV72/p2 by the use of nanobody-aided crystallization. SbsB consists of a seven-domain protein, formed by an amino-terminal cell-wall attachment domain and six consecutive immunoglobulin-like domains, that organize into a ϕ -shaped disk-like monomeric crystallization unit stabilized by interdomain Ca^{2+} ion coordination. A Ca^{2+} -dependent switch to the condensed SbsB quaternary structure pre-positions intermolecular contact zones and renders the protein competent for S-layer assembly. On the basis of crystal packing, chemical crosslinking data and cryo-electron microscopy projections, we present a model for the molecular organization of this SLP into a porous protein sheet inside the S-layer. The SbsB lattice represents a previously undescribed structural model for protein assemblies and may advance our understanding of SLP physiology and self-assembly, as well as the rational design of engineered higher-order structures for biotechnology^{4–7}.

SLPs (40–200 kDa) self-assemble into paracrystalline lattices that are non-covalently attached to the cell surface, usually by means of one or more N-terminal glycan-binding domains^{1,3,11}. S-layers form a continuous outer protein layer, or exoskeleton, that serves as a protective and/or scaffolding layer, a molecular sieve for nutrient uptake, and a contact zone with the extracellular environment, for example in promoting cell adhesion in pathogenic bacteria such as *Clostridium difficile* and *Bacillus anthracis*^{1,12,13}.

Here we use nanobodies as crystallization chaperones¹⁴ to break the intrinsic self-polymerizing propensity into two-dimensional lattices of the SLP SbsB. We report the 2.4-Å X-ray structure of full-length, mature SbsB (SbsB_{32–920}; residues 32–920) from *Geobacillus stearothermophilus* PV72/p2 (ref. 15), a biophysically¹⁶, biochemically^{17,18} and biotechnologically^{4,19,20} well-characterized SLP (Fig. 1). SbsB is a 98-kDa polypeptide that can be functionally divided into a cell-wall-anchoring N terminus (residues 32–208, consisting of three S-layer homology (SLH) motifs¹⁶) and a carboxy-terminal crystallization region that naturally assembles in an oblique (P1) two-dimensional lattice with the unit-cell vectors $\mathbf{a} = 104$ Å, $\mathbf{b} = 79$ Å and base angle $\gamma = 81^\circ$ (ref. 19). Nanobodies were raised against an assembly-incompetent SbsB mutant caused by insertion of the haemagglutinin (HA) tag at residue 744 (SbsB_{T744HA})¹⁸. Both the mutant and wild-type SbsB (Supplementary Table 1) were crystallized in complex with nanobody KB6 (NbKB6; Supplementary Fig. 1). The SbsB crystal structure reveals an extended molecule built from

seven distinct domains (domains I–VII; Fig. 1a). The N-terminal 169 amino-acid residues of the mature protein (domain I; residues 32–201), corresponding to the cell-wall attachment domain²¹, are found unresolved in the electron density maps. Domains II (residues 202–292) and III (residues 296–385) are two consecutive C₁-type immunoglobulin folds. Domains IV (residues 391–499), V (residues 502–625) and VI (residues 631–737) are consecutive I-type immunoglobulin folds, and domain VII (residues 738–920) comprises a mixed fold with an immunoglobulin-like β -sandwich core and a random-coil sub-domain (residues 769–849) inserted between strands B and C (Fig. 1a and Supplementary Fig. 2).

In the crystal structure, domains II–VII arrange into a disk-like ϕ -shaped quaternary structure 70 Å by 110 Å wide, and 35 Å thick (Fig. 1a, b). In this arrangement, domains IV–VII lie in a single plane and form an annular structure that encloses a central cavity about 24 Å in diameter. This cavity is transversed by a nine-residue linker between domains IV and III, which is positioned on the rim of the ring, underneath domain VI (Fig. 1b). Finally, domain II protrudes from the ring structure, giving shape to the 'leg' of the ϕ -shaped monomeric unit (Fig. 1a, b). The ϕ -shaped conformation seen in the crystals is also found for the monomeric protein in solution, as shown by overlays with *ab initio* shapes reconstructed from small-angle X-ray scattering (SAXS) curves obtained for the SbsB_{32–920}, T744HA:NbKB6 and SbsB_{209–920}, T744HA:NbKB6 complexes (an N-terminal deletion mutant removing the unresolved domain I) (Fig. 2 and Supplementary Fig. 3). The averaged *ab initio* shapes for SbsB_{32–920}, T744HA:NbKB6 show an additional density near domain II, corresponding to the N-terminal attachment domain, which was found disordered in the X-ray structure. Comparison of the experimental scattering data with calculated curves from molecular models of SbsB_{32–920}, T744HA:NbKB6 with domain I either unstructured or modelled with the *B. anthracis* Sap SLH domain^{10,13} gives fitting χ^2 values of 2.4 and 13.3, respectively (Supplementary Fig. 3), indicating that the N-terminal domain is largely unstructured in the non-polymerized SbsB and/or in the absence of secondary cell wall polymer (SCWP) (Supplementary Fig. 3 and Methods).

The inter-domain and intra-domain contacts in the SbsB quaternary structure are mediated by Ca^{2+} ions (Fig. 1). A first Ca^{2+} (Ca-1) is shared between domains IV and VII and octahedrally coordinated by residues coming from both domains (Fig. 1 and Supplementary Fig. 4). A second Ca^{2+} site (Ca-2; Fig. 1 and Supplementary Fig. 4) is located at the hinge between domains V and VI. A third site, Ca-3, is positioned in an extended loop of domain VI, near the contact zone with domain VII (Fig. 1 and Supplementary Fig. 4). The fourth site, Ca-4, forms an intra-domain bridge in a coil region of domain VII (Fig. 1 and Supplementary Fig. 4). These Ca^{2+} sites assist in the maintenance of inter-domain contacts and thus probably represent important determinants for the SbsB quaternary structure. Chemically denatured SbsB readily refolds and assembles into S-layers on removal of the chaotropic agents, but

¹Structural and Molecular Microbiology, VIB Department of Structural Biology, VIB, Pleinlaan 2, 1050 Brussels, Belgium. ²Structural Biology Brussels, Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels, Belgium. ³Unité G5, Biologie structurale de la sécrétion bactérienne, Institut Pasteur, 25–28 rue du Docteur Roux, 75015 Paris, France. ⁴Unité de recherche mixte 3538, Centre national de la recherche scientifique, Institut Pasteur, 25–28 rue du Docteur Roux, 75015 Paris, France. ⁵Department of Chemistry, Institute of Structural and Molecular Biology, University College London, London WC1H 0AJ, UK.

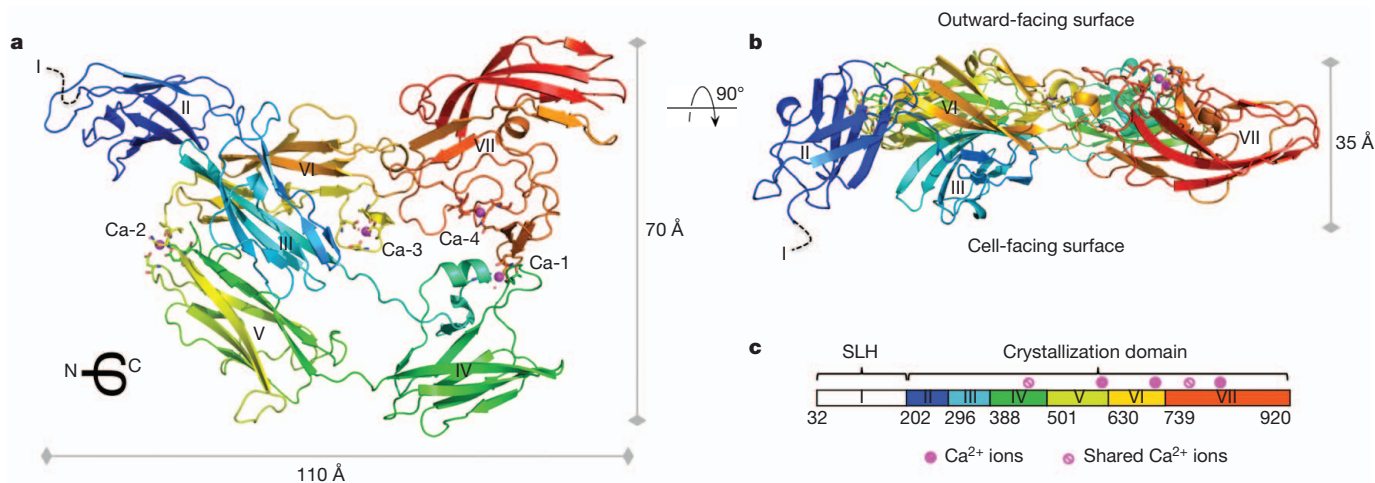


Figure 1 | X-ray structure of *G. stearotherophilus* SbsB. Ribbon representations of the SbsB monomer. **a**, Viewed from the cell-facing side. **(b)** Viewed from inside the plane of the S-layer. SbsB is coloured from blue to red from N terminus to C terminus; Ca^{2+} ions (magenta) and Ca^{2+} -

only in the presence of Ca^{2+} , a characteristic that is shared by many SLPs^{16,22}. Our SAXS data show that EDTA-treated SbsB samples lose quaternary structure (Fig. 2), but the individual domains retain their native secondary structure content as judged by circular dichroism spectra (Fig. 2 inset; thermal denaturation temperatures were 75 and 69 °C in the absence and presence of EDTA, respectively (data not shown)). Thus, Ca^{2+} binding in SbsB triggers a structural transition from an extended chain of folded immunoglobulin domains into a ϕ -shaped quaternary conformation that renders the protein competent for assembly (isothermal titration calorimetry of EDTA or Ca^{2+} titrations reveal Ca^{2+} dissociation constants for the four sites ranging from 0.9 to 110.0 μM ; data not shown). Inter-domain Ca^{2+} binding between neighbouring immunoglobulin domains has previously been observed in cadherins²³. Analogously to SbsB, Ca^{2+} binding in C-cadherin switches the protein structure from a loose rope of consecutive immunoglobulin domains to an active, rigidified hook-like architecture²⁴.

To gain insight into SbsB polymerization, cryo-electron microscopy (cryo-EM) images of SbsB_{209–920} lattices were obtained and combined with the X-ray structure to yield a molecular model of the S-layer. In cryo-EM, the *in vitro* self-assembled lattice of SbsB_{209–920} (Fig. 3a and Supplementary Fig. 5) has unit-cell dimensions $\mathbf{a}^{\text{S-layer}} = 99 \text{ Å}$, $\mathbf{b}^{\text{S-layer}} = 76 \text{ Å}$ and base angle $\gamma = 81^\circ$, which are close to the literature values for full-length SbsB from negative-stain electron microscopy ($\mathbf{a} = 104 \text{ Å}$, $\mathbf{b} = 79 \text{ Å}$, base angle $\gamma = 81^\circ$)¹⁹. The three-dimensional crystal packing of the SbsB_{32–920}:NbKB6 complex features stacked planes of SbsB layers with pseudotranslational vectors $\mathbf{a}^{\text{X-ray}} = 115.4 \text{ Å}$, $\mathbf{b}^{\text{X-ray}} = 75.2 \text{ Å}$ and base angle $\gamma = 72^\circ$, very similar to the S-layer lattice seen in cryo-EM (Fig. 3b). In particular, $\mathbf{b}^{\text{X-ray}}$ seems isomorphous with $\mathbf{b}^{\text{S-layer}}$, whereas $\mathbf{a}^{\text{X-ray}}$ is elongated by 17 Å and the base angle is shifted by 7° as a result of the inclusion of the nanobody between SbsB monomers (Fig. 3b). We exploited this fortuitous finding of isosymmetry to derive a model for the two-dimensional SbsB lattice in the S-layer. Replacing $\mathbf{a}^{\text{X-ray}}$ with $\mathbf{a}^{\text{S-layer}}$ results in a continuous SbsB lattice, which is maintained through two discrete intermolecular contact zones (Fig. 3c and Supplementary Fig. 6). Along the \mathbf{b} -vector, monomers contact through domains IV and VII, covering a total surface area of 450 Å², and in agreement with published crosslinking screens¹⁷ (Fig. 3c and Supplementary Fig. 6). Along the calculated \mathbf{a} -vector, the contact is formed through domain II. Overlays with cryo-EM projection maps show the close match of the calculated S-layer model (Fig. 3c and Supplementary Figs 7 and 8). A local discrepancy is seen at domain II, where the projection maps show a displacement of the corresponding density closer towards domain IV

coordinating residues are shown as spheres and sticks, respectively. For clarity, nanobody NbKB6 is not shown. **c**, Schematic representation of the mature SbsB; domains II–VII are coloured as in **a**. Residue numbers indicate domain borders. Magenta dots indicate the locations of Ca^{2+} sites.

of the adjacent molecule (IV^c; Fig. 3c and Supplementary Fig. 8). Molecular dynamics simulations employing intermolecular distance restraints derived from photochemical crosslinking experiments (Fig. 3c and Supplementary Fig. 9) settle domain II in a conformation that coincides with the density observed in the electron microscopy projection maps (Fig. 3c and Supplementary Fig. 8). The modelled position of domain II encompasses a 490 Å² contact surface with the adjacent domain IV and was independently confirmed through Cys–Cys crosslinking of selectively introduced Cys residues in domains II and IV (Fig. 3c and Supplementary Fig. 10). The repositioning of domain II is made possible through a conformational hinge (residues 292–296) with domain III. Between different cryoEM sections, varying levels of electron density are observed for domain II, indicating

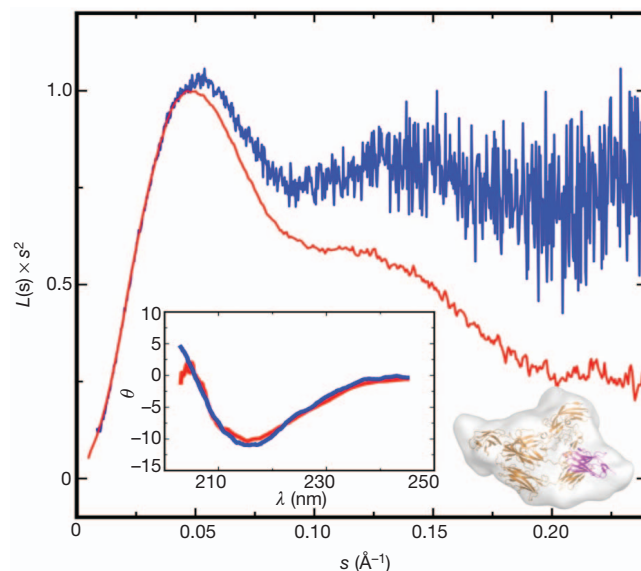


Figure 2 | Solution studies of SbsB quaternary structure. Kratky plots of SbsB_{209–920}:T744HA:NbKB6 before (red) and after the addition of 50 mM EDTA (blue). The native SbsB quaternary structure (averaged *ab initio* shape reconstruction volume is shown in the lower right inset, with SbsB_{209–920}:T744HA and NbKB6 in gold and magenta ribbon, respectively; see Supplementary Fig. 3) is disrupted on the loss of Ca^{2+} , leading to an extended structure consisting of a string of domains II–VII. Under the same conditions, circular dichroism spectra (inset; ellipticity θ is given in units of $10^{-4} \text{ deg cm}^2 \text{ mol}^{-1}$) show that the extended, Ca^{2+} -free protein retains the typical β -sheet conformation, indicating that the individual domains remain folded.

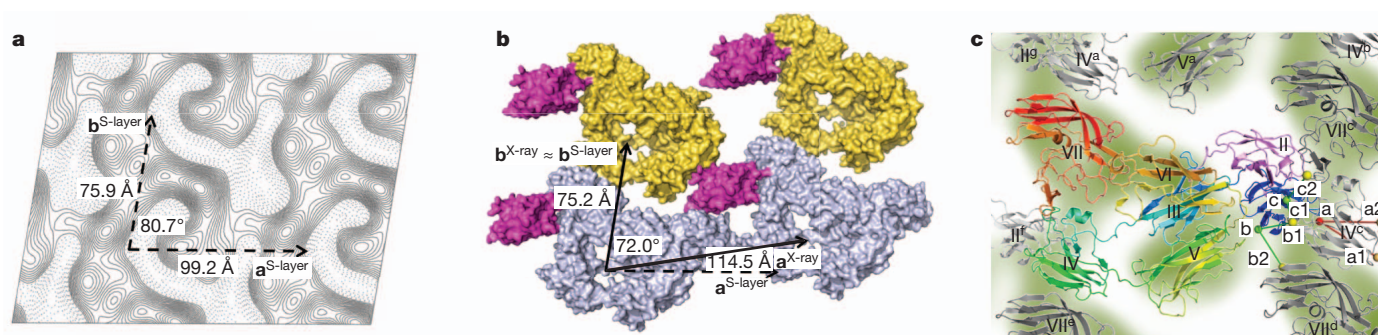


Figure 3 | Cryo-EM imaging and lattice model for the SbsB S-layer.

a, Projection map of *in vitro* grown S-layers of SbsB₂₀₉₋₉₂₀ calculated from a representative image (using a 18 Å resolution cutoff and a 27.0×10^6 Å² surface area, corresponding to 3,600 unit cells; Methods and Supplementary Figs 7 and 8; the unit cell is shown as dashed vectors). **b**, Surface representation of SbsB₃₂₋₉₂₀:NbKB6 three-dimensional crystal packing, showing pseudosymmetrical vectors $\mathbf{a}^{\text{X-ray}}$ and $\mathbf{b}^{\text{X-ray}}$, highly similar to $\mathbf{a}^{\text{S-layer}}$ and $\mathbf{b}^{\text{S-layer}}$ (dashed vectors): $\mathbf{b}^{\text{X-ray}}$ and $\mathbf{b}^{\text{S-layer}}$ are isomorphous, whereas the inclusion of NbKB6 (magenta) results in a mismatch between $\mathbf{a}^{\text{X-ray}}$ and $\mathbf{a}^{\text{S-layer}}$. The figure shows two asymmetric units, each containing two SbsB₃₂₋₉₂₀:NbKB6 complexes (orange and blue, with domains II–VII and III–VII ordered, respectively). **c**, Overlay of

cryo-EM projection map (green) with SbsB S-layer model. SbsB₃₂₋₉₂₀ is coloured as in Fig. 1 or in greyscale (light to dark from N terminus to C terminus) for neighbouring protomers (indicated with superscripts a to g). A local mismatch at domain II is resolved by molecular dynamics modelling (X-ray and modelled conformation coloured violet and blue, respectively) using intermolecular distance restraints derived from crosslinking experiments that show that T281C (labelled a) lies within 22 Å of K498 or K499 (red lines labelled a1 and a2). Intermolecular Cys–Cys crosslinking (green lines) of Cys double mutants T240C–K486C (b–b1), T240C–T862C (b–b2), T268C–K486C (c–c1) and T268C–E780C (c–c2) confirms the modelled position for domain II.

that domain II may not adopt a single, uniform conformation throughout the S-layer (Supplementary Fig. 8). This conformational hinge may allow the S-layer to accommodate local topological defects on the bacterial cell surface.

The sidedness with which the modelled S-layer binds the bacterial cell wall (Figs 3 and 4) is assigned by alignment of the lattice vectors with cell-bound or sacculus-bound SbsB lattices¹⁹ and is in excellent

agreement with a previously published accessibility study (Supplementary Fig. 11)^{17,25}. In this topology, domain II is oriented such that the cell-wall-binding domain I is located on the cellular side of the lattice, poised to bind a peptidoglycan-tethered secondary cell-wall polymer²¹. S-layers form a continuous protein sheet wrapping their bacterial cells. Nutrient uptake and cellular secretions require this sheet to be semipermeable. Cryo-EM projections and the modelled SbsB S-layer show zones of perforation in the protein sheet. An intermolecular cavity 30 Å in average diameter and 70 Å in maximum width is formed, lined by domains II, V, VI and VII (Fig. 4).

The bacterial SbsB lattice represents a previously undescribed structural model for two-dimensional protein assemblies. Large spherical or curved assemblies such as viral coats or bacterial polyhedral microcompartments are usually composed of oligomeric protein units that can be structurally tightly integrated^{26–28}. For example, the hexamer of the CA capsid protein from HIV-1 is held together by α -helical bundles as well as by intermolecular donor-helix exchange²⁹. The oligomers facilitate lattice growth by pre-orienting key assembly residues within the oligomeric scaffold. The SbsB lattice, by contrast, is of P1 symmetry and features a monomer as its morphological unit. Two-dimensional lattice formation follows a two-step process with a Ca²⁺-triggered folding-back of the elongated protein to form a condensed looped quaternary structure. This looped structure prepositions key sites for assembly and renders the protein assembly competent (Fig. 4c). It is likely that the architectural design features of SbsB are applicable to many other SLPs. P1 S-layer lattices and the requirement of Ca²⁺ for assembly are widespread²², and at least for SbsC (17% sequence identity with SbsB) the available structural data show a stretch of multiple immunoglobulin-like domains⁹ (Supplementary Fig. 12). We envisage that the structural insights and the Ca²⁺-dependent conformational switch in S-layer lattice assembly will provide opportunities for the rational design of dynamic, functionalized coating biomaterials.

METHODS SUMMARY

A set of eight unique SbsB-binding nanobodies, corresponding to three VHH families based on sequence diversity in the CDR2 and CDR3 regions (Supplementary Fig. 1), was generated from a llama (*Llama glama*) immunized with a recombinant SbsB₂₀₉₋₉₂₀:T744HA. Recombinant SbsB₂₀₉₋₉₂₀:T744HA and SbsB₃₂₋₉₂₀ were crystallized with a stoichiometric amount of nanobody NbKB6 and their structures were determined in a single-anomalous-wavelength experiment on a SbsB₃₂₋₉₂₀:T744HA:NbKB6 crystal soaked with 2.5 mM GdCl₃ (Supplementary Table 1). The structure for SbsB₃₂₋₉₂₀:NbKB6 was solved by molecular

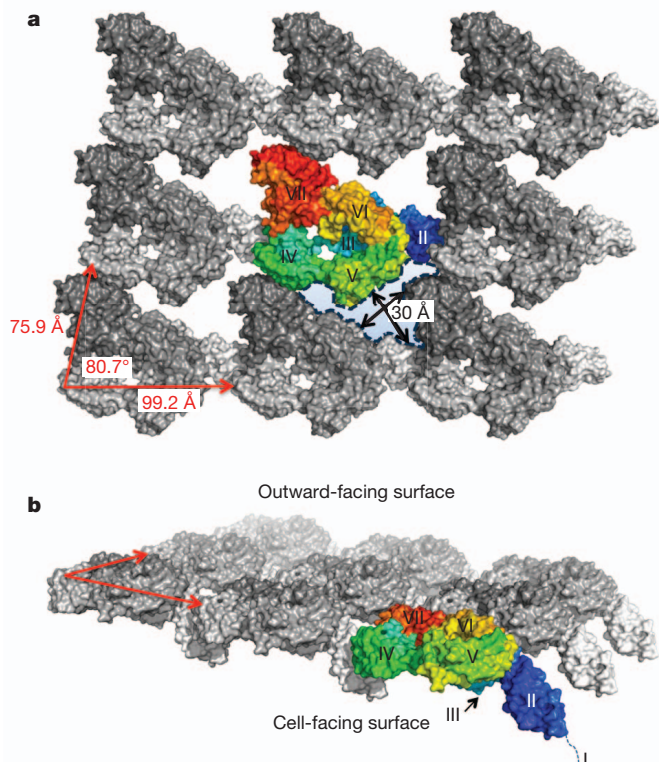


Figure 4 | Model for the *G. stearotherophilus* SbsB S-layer. Surface representation of the modelled SbsB S-layer. **a**, Viewed from the extracellular side. **b**, Tilted in side view. Monomers are coloured in greyscale (light to dark from N terminus to C terminus). For clarity, one monomer is coloured as in Fig. 1. The schematic outline highlights the intermolecular pore formed by three neighbouring subunits. Domain annotation is as in Fig. 1.

substitution and refined to 2.4 Å resolution, providing a model with *R*-factor and free *R*-factor of 18.6% and 23.8%, respectively, and 98% residues in allowed Ramachandran geometry. Circular dichroism spectra were collected at 20 °C on a Jasco J-715 spectropolarimeter, using 200 µl of protein at 0.5 mg ml⁻¹ concentration in buffer A (50 mM Tris-HCl pH 8.0, 150 mM NaCl), with and without 50 mM EDTA. To produce *in vitro* S-layers, purified SbsB_{209–920} at 0.1 mg ml⁻¹ in buffer A supplemented with 2 mM CaCl₂ was concentrated to 1 mg ml⁻¹ and left at 18 °C for a week. Crystals were deposited on glow-discharged carbon-coated 300-mesh grids, flash-frozen in liquid ethane and observed with a Jeol 2010F electron microscope operated at 200 kV. To map residues at the subunit interfaces, SbsB_{T281C} was modified with the photoactivatable and biotin-tagged hetero-bifunctional crosslinker Mts-Atf-LC-biotin, assembled into lattices, and exposed to ultraviolet. Protein dimers were isolated by gel electrophoresis and then digested proteolytically with trypsin. Crosslinked peptide fragments carrying the biotin tag were isolated by affinity purification and subjected to mass spectrometry to determine the peptide mass and sequence (Supplementary Fig. 9).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 21 December 2011; accepted 19 April 2012.

Published online 10 June 2012.

- Sara, M. & Sleytr, U. B. S-layer proteins. *J. Bacteriol.* **182**, 859–868 (2000).
- Glauert, A. M. The fine structure of bacteria. *Br. Med. Bull.* **18**, 245–250 (1962).
- Messner, P., Schäffer, C., Egelseer, E. M. & Sleytr, U. B. in *Prokaryotic Cell Wall Compounds—Structure and Biochemistry* (eds König, H., Claus, H. & Varma, A.) 53–109 (Springer, 2010).
- Sleytr, U. B. *et al.* in *Progress in Molecular Biology and Translational Science* vol. 103 (ed. Howorka, S.) 277–352 (Academic, 2011).
- Howorka, S. Rationally engineering natural protein assemblies in nanobiotechnology. *Curr. Opin. Biotechnol.* **22**, 485–491 (2011).
- Mark, S. S. *et al.* Bionanofabrication of metallic and semiconductor nanoparticle arrays using S-layer protein lattices with different lateral spacings and geometries. *Langmuir* **22**, 3763–3774 (2006).
- Nam, Y. S. *et al.* Biologically templated photocatalytic nanostructures for sustained light-driven water oxidation. *Nature Nanotechnol.* **5**, 340–344 (2010).
- Fagan, R. P. *et al.* Structural insights into the molecular organization of the S-layer from *Clostridium difficile*. *Mol. Microbiol.* **71**, 1308–1322 (2009).
- Pavkov, T. *et al.* The structure and binding behavior of the bacterial cell surface layer protein SbsC. *Structure* **16**, 1226–1237 (2008).
- Kern, J. *et al.* Structure of surface layer homology (SLH) domains from *Bacillus anthracis* surface array protein. *J. Biol. Chem.* **286**, 26042–26049 (2011).
- Engelhardt, H. & Peters, J. Structural research on surface layers: a focus on stability, surface layer homology domains, and surface layer–cell wall interactions. *J. Struct. Biol.* **124**, 276–302 (1998).
- Calabi, E., Calabi, F., Phillips, A. D. & Fairweather, N. F. Binding of *Clostridium difficile* surface layer proteins to gastrointestinal tissues. *Infect. Immun.* **70**, 5770–5778 (2002).
- Kern, J. & Schneewind, O. BslA, the S-layer adhesin of *B. anthracis*, is a virulence factor for anthrax pathogenesis. *Mol. Microbiol.* **75**, 324–332 (2010).
- Steyaert, J. & Kobilka, B. K. Nanobody stabilization of G protein-coupled receptor conformational states. *Curr. Opin. Struct. Biol.* **21**, 567–572 (2011).
- Sara, M. *et al.* Dynamics in oxygen-induced changes in S-layer protein synthesis from *Bacillus stearothermophilus* PV72 and the S-layer-deficient variant T5 in continuous culture and studies of the cell wall composition. *J. Bacteriol.* **178**, 2108–2117 (1996).
- Runzler, D., Huber, C., Moll, D., Kohler, G. & Sara, M. Biophysical characterization of the entire bacterial surface layer protein SbsB and its two distinct functional domains. *J. Biol. Chem.* **279**, 5207–5215 (2004).
- Kinns, H. & Howorka, S. The surface location of individual residues in a bacterial S-layer protein. *J. Mol. Biol.* **377**, 589–604 (2008).
- Kinns, H., Badelt-Lichtblau, H., Egelseer, E. M., Sleytr, U. B. & Howorka, S. Identifying assembly-inhibiting and assembly-tolerant sites in the SbsB S-layer protein from *Geobacillus stearothermophilus*. *J. Mol. Biol.* **395**, 742–753 (2010).
- Moll, D. *et al.* S-layer–streptavidin fusion proteins as template for nanopatterned molecular arrays. *Proc. Natl Acad. Sci. USA* **99**, 14646–14651 (2002).
- Mader, C., Kupcu, S., Sara, M. & Sleytr, U. B. Stabilizing effect of an S-layer on liposomes towards thermal or mechanical stress. *Biochim. Biophys. Acta* **1418**, 106–116 (1999).
- Sara, M., Egelseer, E. M., Dekitsch, C. & Sleytr, U. B. Identification of two binding domains, one for peptidoglycan and another for a secondary cell wall polymer, on the N-terminal part of the S-layer protein SbsB from *Bacillus stearothermophilus* PV72/p2. *J. Bacteriol.* **180**, 6780–6783 (1998).
- Pavkov-Keller, T., Howorka, S. & Keller, W. in *Progress in Molecular Biology and Translational Science* vol. 103 (ed. Howorka, S.) 73–130 (Academic, 2011).
- Boggon, T. J. *et al.* C-cadherin ectodomain structure and implications for cell adhesion mechanisms. *Science* **296**, 1308–1313 (2002).
- Sotomayor, M. & Schulten, K. The allosteric role of the Ca²⁺ switch in adhesion and elasticity of C-cadherin. *Biophys. J.* **94**, 4621–4633 (2008).
- Howorka, S. *et al.* Surface-accessible residues in the monomeric and assembled forms of a bacterial surface layer protein. *J. Biol. Chem.* **275**, 37876–37886 (2000).
- Li, L., Jose, J., Xiang, Y., Kuhn, R. J. & Rossmann, M. G. Structural changes of envelope proteins during alphavirus fusion. *Nature* **468**, 705–708 (2010).
- Voss, J. E. *et al.* Glycoprotein organization of Chikungunya virus particles revealed by X-ray crystallography. *Nature* **468**, 709–712 (2010).
- Yeates, T. O., Thompson, M. C. & Bobik, T. A. The protein shells of bacterial microcompartment organelles. *Curr. Opin. Struct. Biol.* **21**, 223–231 (2011).
- Ganser-Pornillos, B. K., Cheng, A. & Yeager, M. Structure of full-length HIV-1CA: a model for the mature capsid lattice. *Cell* **131**, 70–79 (2007).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank staff at beamlines Proxima-1 and SWING (Soleil), beamline BM30A (European Synchrotron Radiation Facility) and beamline X33 (Deutsches Elektronen-Synchrotron) for technical assistance during data collection. We thank G. Waksman for suggestions on the manuscript, and D. Levy for advice on cryo-EM grid preparation and data collection. This work was supported by VIB (Vlaams Instituut voor Biotechnologie) grant PRJ9 to H.R., the Fonds Wetenschappelijk Onderzoek-Vlaanderen through an Odysseus grant to H.R., a postdoctoral fellowship to A.P.G. and research grant FWO551 to J.S. R.F. is supported by the Centre National de la Recherche Scientifique (CNRS) and the Institut Pasteur. G.P.A. is supported by the CNRS. E.P. received support from Interuniversity Attraction Poles grant P6/19. D.P. was funded by Biotechnology and Biological Sciences Research Council grant BB/E010466/1, and S.H. is grateful for support from University College London. We acknowledge Professor Emeritus U. Sleytr for his lifelong devotion to S-layer protein research and G. *stearothermophilus* SbsB in particular.

Author Contributions K.B. produced and crystallized SbsB:NbKB6 complexes, collected and analysed diffraction and SAXS data and determined their structures. A.P.G. analysed SAXS data and collected SAXS, isothermal titration calorimetry and circular dichroism data. N.V.G. performed mutagenesis and crosslinking experiments. R.F. grew S-layers *in vitro* and analysed cryo-EM data. G.P.A. prepared cryo-EM grids and collected cryo-EM data. E.P. and J.S. produced SbsB-binding nanobodies. D.P. performed crosslinking experiments. S.H. produced SbsB constructs, supervised D.P. and wrote the manuscript. H.R. supervised the work, collected diffraction data, solved and analysed the structures and wrote the manuscript.

Author Information Coordinates and structure factors for SbsB_{32–920}:NbKB6 are deposited in the Protein Data Bank under accession code 4AQ1. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to H.R. (han.remaut@vib-vub.be).

METHODS

Expression and purification. Nanobodies were expressed as C-terminal His₆ fusions in *Escherichia coli* WK6 periplasm, using the cAb-Lys3-pHEN6 vector³⁰. SbsB_{209–920}, T744HA (containing residues 209–920 of the mature protein), SbsB_{T744HA} and wild-type SbsB were expressed in *E. coli* B834 (DE3) pLysS cytoplasm as fusion products carrying a C-terminal His₆ tag²⁵. Cells expressing NbKB6, grown overnight at 20 °C in lysogeny broth, were washed and incubated for 20 min at 4 °C with the buffer 20 mM Tris/HCl pH 8.0, 20% sucrose, 5 mM EDTA, 0.1 mg ml^{−1} lysozyme to make spheroplasts. Periplasmic proteins were isolated from spheroplasts by centrifugation (20,000g, 20 min) and dialysed against buffer A (50 mM Tris-HCl pH 8.0, 150 mM NaCl). Pellets from SbsB-expressing cultures were resuspended in buffer A, sonicated and centrifuged for 20 min at 20,000g. Supernatants were mixed with the NbKB6 periplasmic extract, and SbsB:NbKB6 complexes were purified by metal-chelating affinity chromatography employing an imidazole gradient from 40 to 400 mM in buffer A. Remaining protein contaminants and excess NbKB6 were removed by ion-exchange chromatography (HiTrap Q FF; GE Healthcare; buffer: 20 mM Tris-HCl pH 8.0, 40–500 mM NaCl) and hydrophobic interaction chromatography (HiTrap Phenyl HP; GE Healthcare; buffer: 20 mM Tris-HCl pH 8.0, 10 mM NaCl, 1–0 M (NH₄)₂SO₄). No additional Ca²⁺ was added during purification. Our unpublished isothermal titration calorimetry data show that the four Ca²⁺ ions in SbsB have dissociation constants ranging from 0.9 to 110.0 μM. According to ref. 31 the *E. coli* cytoplasm is maintained at 90 ± 10 mM free Ca²⁺, irrespective of extracellular Ca²⁺ levels. Thus, we believe that the protein used for our studies bound Ca²⁺ during cell lysis, when it came in contact with the cellular debris and the pool of cell wall-bound Ca²⁺, which was estimated to be in the range 10–100 mM.

Crystallization and data collection. On purification, wild-type SbsB_{32–920} or SbsB_{209–920} readily aggregates into high-molecular-mass species. To overcome the intrinsic propensity for aggregation or polymerization, wild-type SbsB_{32–920} or SbsB_{209–920} was purified in complex with NbKB6. Purified SbsB:NbKB6 complexes show monodisperse binary complexes that remain stable for days at room temperature and are amenable to three-dimensional crystallization. The role of NbKB6 as a crystallization chaperone is dual. Besides its role in maintaining SbsB in a monodisperse form before crystallization, the nanobody further aids crystallization by providing a contact surface between stacked protein layers inside the three-dimensional crystals (Supplementary Figs 13 and 14). Rod-like crystals (0.5 × 0.07 × 0.05 mm³) of SbsB:NbKB6 complexes (at 9.5 mg ml^{−1} in 10 mM Tris-HCl pH 8.0, 10 mM NaCl) were obtained by vapour diffusion using 0.2 M potassium isothiocyanate, 0.1 M Bis-Tris propane pH 6.5 and 20% poly(ethylene glycol) 3350. Crystals (in crystallization solution supplemented with 10% (v/v) glycerol) were flash-cooled in liquid nitrogen. Diffraction data for SbsB_{32–920}, T744HA:NbKB6 and SbsB_{32–920}:NbKB6 were collected at 100 K at the PROXIMA-1 beamline at SOLEIL (Saint-Aubin, Evry, France) and beamline BM30A at the European Synchrotron Radiation Facility (Grenoble, France). Data were integrated and scaled using the XDS package³². The structure of SbsB_{32–920}, T744HA:NbKB6 was solved using a GdCl₃-soaked crystal in a single-anomalous-dispersion experiment at 7,249 eV (Supplementary Table 1). An SbsB_{32–920}, T744HA:NbKB6 crystal was soaked for 90 s in crystallization buffer supplemented with 10% (v/v) glycerol and 2.5 mM GdCl₃ before being flash-cooled in liquid nitrogen. Heavy-atom location (26 Gd³⁺ sites per two SbsB molecules in the asymmetric unit), phase calculation, density modification and initial model building were performed using the Phenix program package³³ and resulted in good-quality electron density maps as judged by the figure of merit (0.51) and the successful autotracing of the two SbsB and two NbKB6 copies in the asymmetric unit. The autotraced models were completed by manual model building using Coot³⁴ and refined to an *R*-factor and free *R*-factor of 31.6% and 36.6%, respectively, using REFMAC5 (ref. 35). The final SbsB structure was refined^{34,35} against data from a native SbsB_{32–920}:NbKB6 crystal collected to 2.4 Å resolution (phased by molecular substitution with the SbsB_{32–920}, T744HA:NbKB6 model), providing a model of 18.6% and 23.6% *R*-factor and free *R*-factor, respectively, containing two copies of NbKB6 (chains B and D) and two SbsB molecules encompassing residues 202–920 (chain A) or residues 295–920 (chain C) and having 98% of residues with ϕ and ψ angles in favoured regions of the Ramachandran diagram (Supplementary Table 1).

SAXS. SAXS data for SbsB_{32–920}, T744HA:NbKB6 and SbsB_{209–920}, T744HA:NbKB6 were collected on the SWING beamline at SOLEIL and beamline X33 at Deutsches Elektronen-Synchrotron (Hamburg, Germany), respectively. Concentrated SbsB_{32–920}, T744HA:NbKB6 (25 μl; 36 mg ml^{−1} in buffer A) was injected into a size-exclusion column (Shodex KW402.5) and eluted directly into the SAXS flow-through capillary cell at a flow rate of 150 μl min^{−1}. A total of 250 solution scattering curves (each with a 0.5 s recording time) were measured for the main elution peak corresponding to the monomeric protein and were averaged after buffer subtraction. All SAXS curves were measured at 20 °C over the range of

momentum transfer $0.005 < Q = 4\pi \sin(\theta)/\lambda < 0.60 \text{ Å}^{-1}$, where 2θ is the total scattering angle and $\lambda = 1.5 \text{ Å}$ is the X-ray wavelength. Samples for SbsB_{209–920}, T744HA:NbKB6 were measured at three concentrations (0.5, 3 and 5 mg ml^{−1}) in buffer A, with and without 50 mM EDTA. Data processing and analysis were performed with PRIMUS³⁶. The radius of gyration R_g and distance distribution function $p(r)$ were calculated using GNOM³⁷. CRY SOL³⁸ was run in default mode (using 50 as the number of degrees of freedom) to generate SAXS curves from the atomic coordinates of the crystallographic structures. Three-dimensional envelopes of SbsB_{T744HA}:NbKB6 and SbsB_{209–920}, T744HA:NbKB6 were calculated *ab initio* from the experimental SAXS curves, using a simulated annealing protocol as implemented in DAMMIF³⁹. SUPCOMB⁴⁰ was used to overlay atomic coordinates on *ab initio* reconstructed shapes and estimate the similarity between models. The reconstructed shapes were very reproducible over the 15 calculated models, and superimposed with average normalized spatial discrepancy value below 1.5. The Ensemble Optimization Method (EOM) implemented in the ATSAS³⁵ suite was used to model the relative orientation and composition of the N-terminal domain. For EOM, domain I was either considered unstructured or modelled using the *B. anthracis* Sap SLH domain^{10,13} as a template. In each implementation this domain was arranged in 10,000 random conformations and the rest of the molecule was fixed with a unique conformation. The fitting of the curves calculated from the modelled ensembles (obtained using both strategies) to the experimental scattering curve suggests that domain I is largely disordered. Volumetric maps were computed from the averaged bead model using the SITUS package⁴¹ and displayed using CHIMERA⁴².

Cryo-electron microscopy. Purified SbsB_{209–920} at 0.1 mg ml^{−1} in a buffer containing 50 mM Tris-HCl pH 8.0, 150 mM NaCl and 2 mM CaCl₂ was concentrated to 1 mg ml^{−1} and left at 18 °C for a week. Crystals were then withdrawn and kept at 4 °C. Two-dimensional crystals were adsorbed for 1 min on glow-discharged carbon-coated 300-mesh grids and were flash-frozen in liquid ethane using an automated cryo-plunger (Leica EMPG). The cryo-grids were then transferred to a Jeol 2010F electron microscope operated at 200 kV using a Gatan cryo transfer system (Gatan 626DH). Images were directly assessed using low-electron-dose procedures at a nominal magnification of ×40,000 (real magnification ×54,150) on a 15-μm 4,000 × 4,000-pixel² Gatan charge-coupled device camera corresponding to 2.77 Å per pixel (defocus range 500–1,500 nm).

Cryo-EM images (60 in total) of two-dimensional crystals were treated using the 2dx image processing package⁴³. In brief, the best images (15 in total) were unbent by running two cycles of CCUNBEND. The contrast transfer phase was corrected after the defocus had been determined using CTFFIND⁴⁴. The final map was calculated by merging the best images (five images) using 2dx-merge (ref. 45). Projection maps were calculated using programs of the CCP4 package⁴⁶. An isotropic temperature factor ($B = 500$) was applied to compensate for the degradation of image amplitudes. Density contouring was at 0.25 root mean square with densities above the mean contoured by solid lines. The merged projection map was obtained using 123 spots and gave an IQ (signal-to-noise-ratio-based Index of Quality) weighted phase residual of 28.4 on the whole range of the power spectrum (IQ 1–8). The data were filtered at 18 Å to include only high-IQ spots for calculation of the final map (Supplementary Fig. 7).

Mts-Atf-LC-biotin crosslinking experiments. To confirm the location of domain II, a chemical crosslinking procedure (summarized in Supplementary Fig. 9a) was used to identify peptide sequences that in the assembled S-layer lattice of SbsB are proximal (within 22 Å) to the surface-exposed cysteine residue T281C, introduced in domain II. Cell lysates containing His₆-tagged SbsB_{32–920}, T281C (expressed in *E. coli* JM109(DE3)) were loaded on Ni²⁺-nitrilotriacetate resin (600 μl; Qiagen) in buffer B (50 mM NaH₂PO₄, 300 mM NaCl, 10 mM imidazole) supplemented with 0.5 mM dithiothreitol (DTT) (Supplementary Fig. 9a, step 1) and washed three times with prechilled buffer B (600 μl each, in the absence of DTT) to remove non-His₆-tagged protein. The column-bound protein (about 250 μg) was chemically modified with the thiol-reactive moiety of the heterobifunctional, photoactivatable and biotin-tagged crosslinker Mts-Atf-LC-biotin (Pierce) by a 20-min incubation (under subdued light) at 4 °C in 500 μM crosslinker in buffer B (Supplementary Fig. 9a, step 2). Excess reagent was removed by short centrifugation, followed by four washes with prechilled buffer B (600 μl each). The modified protein was eluted in buffer B supplemented with 250 mM imidazole and mixed with a freshly prepared suspension of peptidoglycan sacculi of *G. stearothermophilus* PV72/p2 (3:1 weight ratio of eluted SbsB protein to peptidoglycan sacculi) and incubated overnight at 4 °C, yielding S-layer assembly products as shown by negative staining and transmission electron microscopy (Supplementary Fig. 9a, step 3). The photosensitive azidotrifluorophenyl-moiety (Atf) of Mts-Atf-LC-biotin-modified SbsB_{32–920}, T281C was activated by exposure to ultraviolet (360 nm for 15 min at a distance of 2–5 cm from the solution) leading to the formation of a covalent bond between the crosslinker and an accessible lysine residue within 22 Å distance range (Supplementary Fig. 9a,

step 4). Unreacted crosslinker was quenched by adding 20 μ l of 1.5 M Tris-HCl pH 8.0. SbsB dimers and monomers, corresponding to intramolecularly and intermolecularly reacted species, respectively, were isolated by SDS-PAGE (7.5% bisacrylamide gels) (Supplementary Fig. 9a, step 5). The Coomassie-blue-stained gel bands were cut out and destained with an aqueous solution containing 5% methanol and 7% acetic acid. After removing the destaining solution, the gel slices were further purified by rinsing with acetonitrile and water, followed by incubation for 30 min with a mixture of freshly prepared 50% $(\text{NH}_4)_2\text{CO}_3$ buffer in acetonitrile at 37 °C under gentle shaking, and dried. SbsB protein was subjected to proteolytic digestion by swelling the dried gel slices in a 10 ng μ l⁻¹ solution of trypsin (40 μ l; NEB; catalogue no. P8101S) in reaction buffer (50 mM Tris-HCl pH 8.0, 20 mM CaCl_2) for 20 min, followed by the addition of reaction buffer to cover the slice and incubation overnight at 25 °C under gentle shaking (Supplementary Fig. 9a, step 6). The digestion reaction was stopped by the addition of phenylmethylsulphonyl fluoride to a final concentration of 1 mM. Crosslinked peptides were purified from the supernatant by incubation for 45 min at room temperature with streptavidin-agarose beads (20 μ l; Pierce) equilibrated in PBS buffer, supplemented with 10 mM DTT to cleave the disulphide bond between the cysteine-bearing peptide and the peptide containing the Atf-reacted Lys (Supplementary Fig. 9a, step 7). The affinity-bound peptides were eluted by the addition of 0.1% trifluoroacetic acid in ultrapure water (20 μ l) followed by centrifugation (Supplementary Fig. 9a, step 8). The sequence identity of the chemically modified peptide was identified using matrix-assisted laser desorption/ionization mass spectrometry (MALDI-MS) using a Waters MALDI microMX instrument in the reflectron positive or linear positive mode, from 1,200 m/z to 5,000 m/z (Supplementary Fig. 9b, c). The MS traces of SbsB₃₂₋₉₂₀, T281C and wild-type SbsB protein were compared, to identify peaks specific for the crosslinked proteins (Supplementary Fig. 9b). In addition, MS traces for monomeric and dimeric protein were compared, to pinpoint unique major signals (Supplementary Fig. 9b). The m/z values of the identified peaks were then corrected by subtracting the mass of the fragment of the chemical crosslinker (847.02 Da) to yield the peptide mass. The corresponding peptide sequence was inferred by matching the mass to the predicted peptide fragments obtained from the MS-Digest server at the University of California, San Francisco (<http://prospector.ucsf.edu/prospector/mshome.htm>) and allowing for adducts of Na^+ . Results are summarized in Supplementary Fig. 9c.

Molecular dynamics simulations. Molecular dynamics simulations were performed with the software Almost (<http://www.open-almost.org>). Starting from the reconstituted S-layer (Fig. 3c), a pair of SbsB₂₀₉₋₉₂₀ molecules corresponding to the intermolecular contact zone along the S-layer *a* axis was selected (representing the interface between domains II and IV^c; Fig. 3c). An ensemble of structures was generated with domain II randomly oriented with respect to the rest of the domains. Short torsion-angle molecular dynamics simulations (100 ps) were run to ensure that the structures did not contain distorted bond angles. We then used a 22 Å distance cutoff between atom SG of residue T281C and the NZ atom of K498 or K499 in domain IV of the adjacent SbsB monomer to select an initial set of structures for molecular dynamics calculations, run using the Amber03 force field⁴⁷ in stages of 500 steps with step sizes of 0.002 ps for a total of about 1.0 ns. An ensemble of the top ten energy-minimized structures shows an average root mean square deviation of 1.1 Å for the equivalent C α atoms in domain II, showing that the modelling strategy resulted in one discrete conformation for domain II. A representative structure from this ensemble was used to calculate the S-layer model shown in Figs 3c and 4, using the vector replacement method described above.

BM(PEG)₃ crosslinking. On the basis of the modelled position of domain II, SbsB₂₀₀₋₉₂₀ double Cys mutants were made that introduced a surface-exposed

Cys on either side of the intermolecular interface. Cys substitutions were chosen to lie within 18 Å of one another such that they could be crosslinked with the thiol-reactive homo-bifunctional crosslinker BM(PEG)₃ (Pierce; BM(PEG)₃ holds two maleimide groups spaced 17.8 Å apart). In this way, four double mutants were generated: T240C/K486 (II/IV^c: 15.9 Å), T240C/T862C (II/VII^c: 16.9 Å), T268C/K486C (II/IV^c: 9.4 Å) and T268C/E780C (II/IV^d: 17.0 Å) (Fig. 3c and Supplementary Fig. 10). His₆-tagged SbsB₂₀₀₋₉₂₀ double mutants were purified by nickel-affinity chromatography. SbsB₂₀₀₋₉₂₀ mutants were eluted in PBS containing 1 M imidazole and 10 mM DTT to avoid Cys oxidation; they were spin-concentrated to 50 μ M, washed twice in PBS, 10 mM DTT, 1 mM CaCl_2 before incubation overnight at 4 °C to allow S-layer formation. S-layers were then separated from soluble protein by centrifugation (20,000g, 15 min), washed twice with PBS to remove DTT and resuspended in PBS to a final concentration of 40 μ M before the addition of threefold molar excess BM(PEG)₃. After incubation for 60 min at room temperature, non-reacted BM(PEG)₃ was quenched by the addition of 50 mM DTT. Non-crosslinked polymers were redissolved by the addition of 20 mM EDTA, 2.5 M urea, before being loaded on SDS-PAGE gels for analysis (Supplementary Fig. 10; loading buffer contained 100 mM 2-mercaptoethanol to reduce oligomers that might have formed as a result of Cys oxidation).

30. De Genst, E. *et al.* Chemical basis for the affinity maturation of a camel single domain antibody. *J. Biol. Chem.* **279**, 53593–53601 (2004).
31. Gangola, P. & Rosen, B. P. Maintenance of intracellular calcium in *Escherichia coli*. *J. Biol. Chem.* **262**, 12570–12574 (1987).
32. Kabsch, W. Integration, scaling, space-group assignment and post-refinement. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 133–144 (2010).
33. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010).
34. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* **60**, 2126–2132 (2004).
35. Murshudov, G. N. *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D Biol. Crystallogr.* **67**, 355–367 (2011).
36. Konarev, P., Volkov, V. V., Sokolova, A. V., Koch, M. H. J. & Svergun, D. I. PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *J. Appl. Cryst.* **36**, 1277–1282 (2003).
37. Svergun, D. I. Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Cryst.* **25**, 495–503 (1992).
38. Svergun, D. I., Barberato, C. & Koch, M. H. J. CRYSOLE—a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J. Appl. Cryst.* **28**, 768–773 (1995).
39. Svergun, D. I., Petoukhov, M. V. & Koch, M. H. Determination of domain structure of proteins from X-ray solution scattering. *Biophys. J.* **80**, 2946–2953 (2001).
40. Kozin, M. B. & Svergun, D. I. Automated matching of high- and low-resolution structural models. *J. Appl. Cryst.* **34**, 33–41 (2001).
41. Wriggers, W. Using Situs for the integration of multi-resolution structures. *Biophys. Rev.* **2**, 21–27 (2010).
42. Pettersen, E. F. *et al.* UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
43. Gipson, B., Zeng, X., Zhang, Z. Y. & Stahlberg, H. 2dx—user-friendly image processing for 2D crystals. *J. Struct. Biol.* **157**, 64–72 (2007).
44. Mindell, J. A. & Grigorieff, N. Accurate determination of local defocus and specimen tilt in electron microscopy. *J. Struct. Biol.* **142**, 334–347 (2003).
45. Gipson, B., Zeng, X. & Stahlberg, H. 2dx_merge: data management and merging for 2D crystal images. *J. Struct. Biol.* **160**, 375–384 (2007).
46. Collaborative Computational Project Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **50**, 760–763 (1994).
47. Duan, Y. *et al.* A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.* **24**, 1999–2012 (2003).

Programmable single-cell mammalian biocomputers

Simon Ausländer¹, David Ausländer¹, Marius Müller¹, Markus Wieland¹ & Martin Fussenegger^{1,2}

Synthetic biology has advanced the design of standardized control devices that program cellular functions and metabolic activities in living organisms¹. Rational interconnection of these synthetic switches resulted in increasingly complex designer networks that execute input-triggered genetic instructions with precision, robustness and computational logic reminiscent of electronic circuits^{2,3}. Using trigger-controlled transcription factors, which independently control gene expression^{4,5}, and RNA-binding proteins that inhibit the translation of transcripts harbouring specific RNA target motifs^{6,7}, we have designed a set of synthetic transcription–translation control devices that could be rewired in a plug-and-play manner. Here we show that these combinatorial circuits integrated a two-molecule input and performed digital computations with NOT, AND, NAND and N-IMPLY expression logic in single mammalian cells. Functional interconnection of two N-IMPLY variants resulted in bitwise intracellular XOR operations, and a combinatorial arrangement of three logic gates enabled independent cells to perform programmable half-subtractor and half-adder calculations. Individual mammalian cells capable of executing basic molecular arithmetic functions isolated or coordinated to metabolic activities in a predictable, precise and robust manner may provide new treatment strategies and bio-electronic interfaces in future gene-based and cell-based therapies.

Today's electronic devices are controlled by binary calculations executed by electrons flowing through metal wires that are interconnected as multi-layered logic gates. Similarly, natural and synthetic gene circuits can be encoded and chemically rewired to integrate environmental and cellular signals and perform logic metabolic computations controlling physiological activities^{2,3,8–10}. With its defining modular design and standardization principles¹, synthetic biology has significantly advanced the construction of genetic circuit components that control transcription or translation in a trigger-inducible manner^{11–13}. The standardized and modular design of these gene switches improved compatibility for the 'plug-and-play' construction of higher-order networks that function in bacterial, yeast and mammalian cells after minor refinements¹². Examples of synthetic networks with similar circuit topology in different species include logic gates^{2,3,9,14,15}, molecular counters⁸, molecular timing devices¹⁶, synthetic eco-sensing and quorum-sensing/hormone systems^{17,18}, band-pass filters^{19,20} and different types of oscillator that program rhythmic transgene expression with tunable frequency and amplitude^{21,22}.

Many of the transcription-control devices used share a common two-component design: a synthetic transcription factor containing a trigger-controlled DNA-binding motif fused to a transactivation domain binds to a specific operator, thereby activating adjacent minimal promoters in an adjustable manner¹². For example, the antibiotic erythromycin fine-tunes the induction of the synthetic promoter P_{ETR2} by the erythromycin-dependent transactivator (ET1)⁵, and the apple metabolite phloretin modulates the activation of the chimeric promoter P_{TtgR1} by the phloretin-dependent transactivator (TtgA1)⁴ (Fig. 1). RNA controllers are proteins or nucleic acids that vary translation when binding to complementary natural or synthetic RNA motifs engineered into transcripts^{6,7,23}. For example, the bacteriophage

MS2 coat protein (MS2) binds to the MS2box RNA motif ($MS2_{box}$)⁶, and the archaeal ribosomal protein L7Ae interacts with its cognate C/Dbox RNA structure (C/D_{box})^{7,24} (Fig. 1). These transcription- and translation-control components were wired into the following genetic switchboard: input molecules (erythromycin or phloretin) abolish the binding of the transcription factors (ET1 or TtgA1), which otherwise transactivate their cognate promoters (P_{ETR2} or P_{TtgR1}) and drive expression of the RNA-binding protein (MS2 or L7Ae) or a transcript encoding the protein-specific RNA box ($MS2_{box}$ or C/D_{box}) and the destabilized fluorescent reporter gene (encoding d2EYFP or dsRed) (Fig. 1). Both transcription units are functionally connected by the RNA-binding proteins (MS2 or L7Ae) which inhibit the translation of transcripts containing the cognate RNA boxes ($MS2_{box}$ or C/D_{box}) (Fig. 1). Combinatorial assembly and permutation of the controller components in this genetic switchboard enabled the design of distinct computational circuits containing up to six different expression units and performing specific digital logic operations. Each of the Boolean circuits integrates a two-molecule input into a digital ON/OFF expression decision, following the processing logic of NOT, NAND and N-IMPLY gates (Fig. 2; see Supplementary Fig. 1 for control experiments showing component leakiness and threshold-defining settings for fluorescence-activated cell sorting (FACS)). The N-IMPLY circuits, which are exclusively induced in the presence of only one specific input molecule, were wired in two different formats. The first was an AANDNOTB circuitry, in which the translation of a phloretin-controlled TtgA1-mediated P_{TtgR1} -driven d2EYFP transcript containing a C/Dbox in its 5' untranslated region (5' UTR) could be inhibited by erythromycin-controlled ET1-mediated P_{ETR2} -driven expression of the C/Dbox-binding protein L7Ae (Fig. 2a). The second format was a

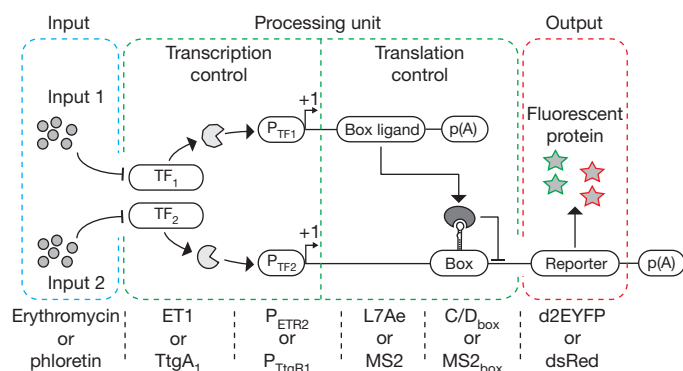


Figure 1 | Genetic switchboard of the biocomputer circuitry. Two input molecules (inputs 1 and 2) inactivate transcription factors (TF₁ and TF₂), which otherwise induce promoters (P_{TF1} and P_{TF2}) that drive the transcription of expression units either encoding an RNA-binding protein (box ligand) or containing an RNA target unit in the 5' UTR (box) and a reporter gene. Both expression units are interconnected through the box ligand, whose interaction with its target box inhibits reporter gene translation. Sequential wiring of the input module (blue) to the processing unit providing bitwise integrations of transcription and translation control activities (green) produces a digital output (red). See Supplementary Information for extended figure legends.

¹Department of Biosystems Science and Engineering, Eidgenössische Technische Hochschule Zürich, Mattenstrasse 26, CH-4058 Basel, Switzerland. ²Faculty of Science, University of Basel, Mattenstrasse 26, CH-4058 Basel, Switzerland.

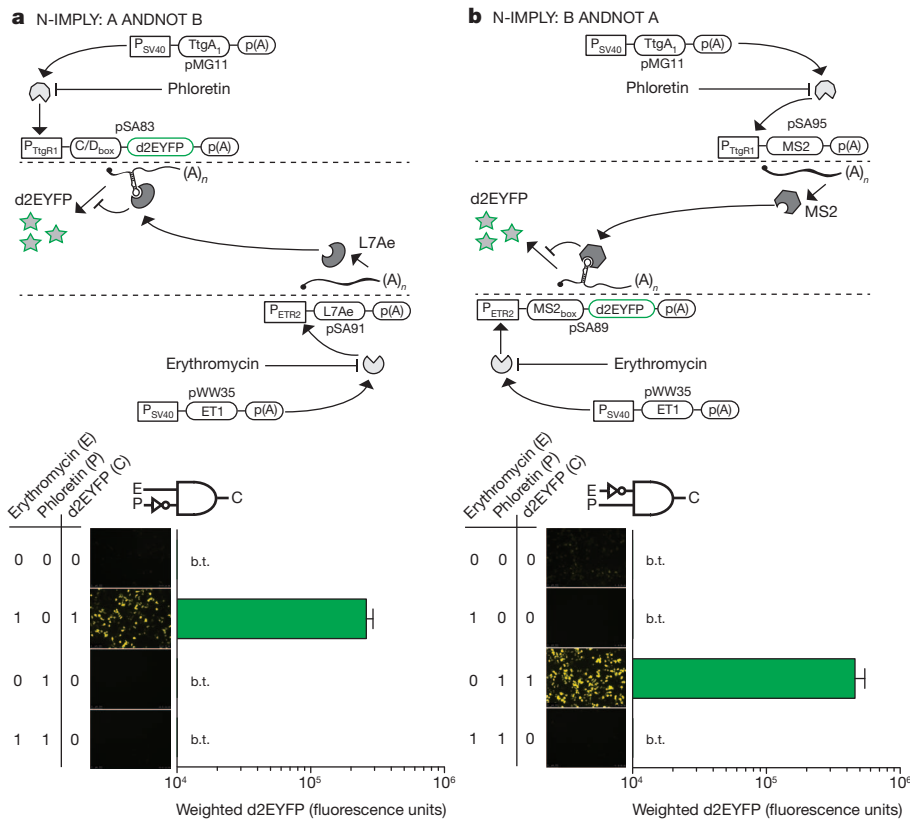
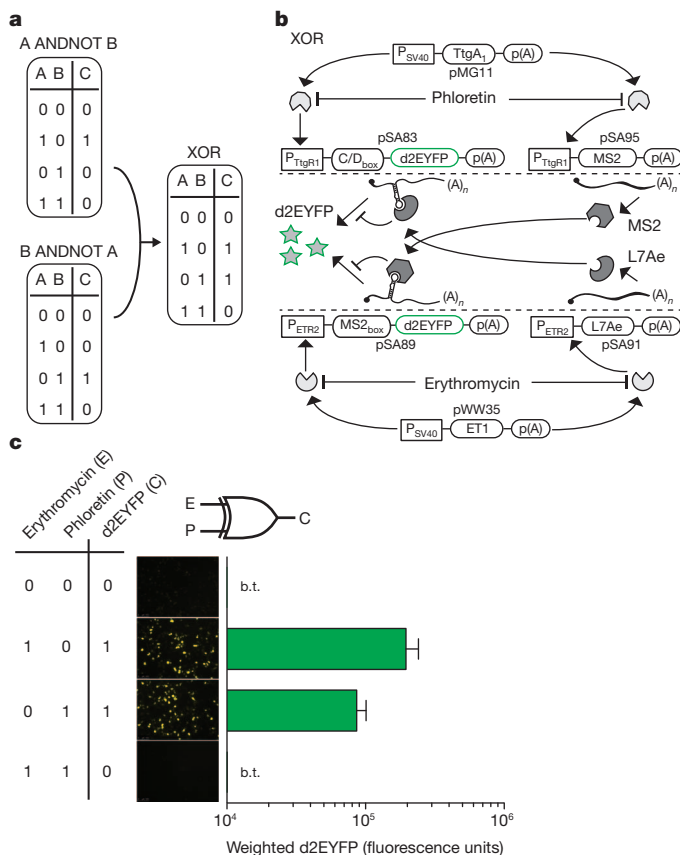


Figure 2 | Design and processing performance of synthetic N-IMPLY gates in human cells. **a**, A ANDNOT B logic gate. By combining the two input signals erythromycin and phloretin in accordance with the truth table, transfected HEK-293 cells are programmed to produce d2EYFP exclusively in the presence of erythromycin and not phloretin as shown by fluorescence microscopy and FACS analysis. **b**, B ANDNOT A logic gate. By combining the two input signals

erythromycin and phloretin in accordance with the truth table, transfected HEK-293 cells are programmed to produce d2EYFP exclusively in the presence of phloretin and not erythromycin as shown by fluorescence microscopy and FACS analysis. b.t., below the threshold of 10⁴ fluorescence units. Error bars represent s.d.; n = 3.



B ANDNOT A circuitry, in which an erythromycin-controlled ET1-mediated d2EYFP transcript containing a MS2box in its 5' UTR could be inhibited by phloretin-controlled P_{TigR1}-driven expression of the MS2 protein (Fig. 2b). After simultaneous transfection of the system-encoding vectors into human embryonic kidney cells (HEK-293), fluorescent reporter gene expression was digitally controlled and exclusively induced when either erythromycin (A ANDNOT B) or phloretin (B ANDNOT A) was present (Fig. 2a, b). As indicated by the consistency of fluorescence microscopy and FACS-mediated single-cell analysis, the cellular N-IMPLY computations executed by the interconnected transcription/translation controllers were remarkably robust (Fig. 2a, b). Detailed control experiments confirmed that the inducer molecules did not influence cell viability^{4,5} and showed that they had no effect on constitutive d2EYFP or dsRed expression levels at repressing concentrations (Supplementary Fig. 2a, b). Furthermore, there was no relevant interference between the RNA-controlling modules, which ensures high specificity of the engineered N-IMPLY gates (Supplementary Fig. 3a, b).

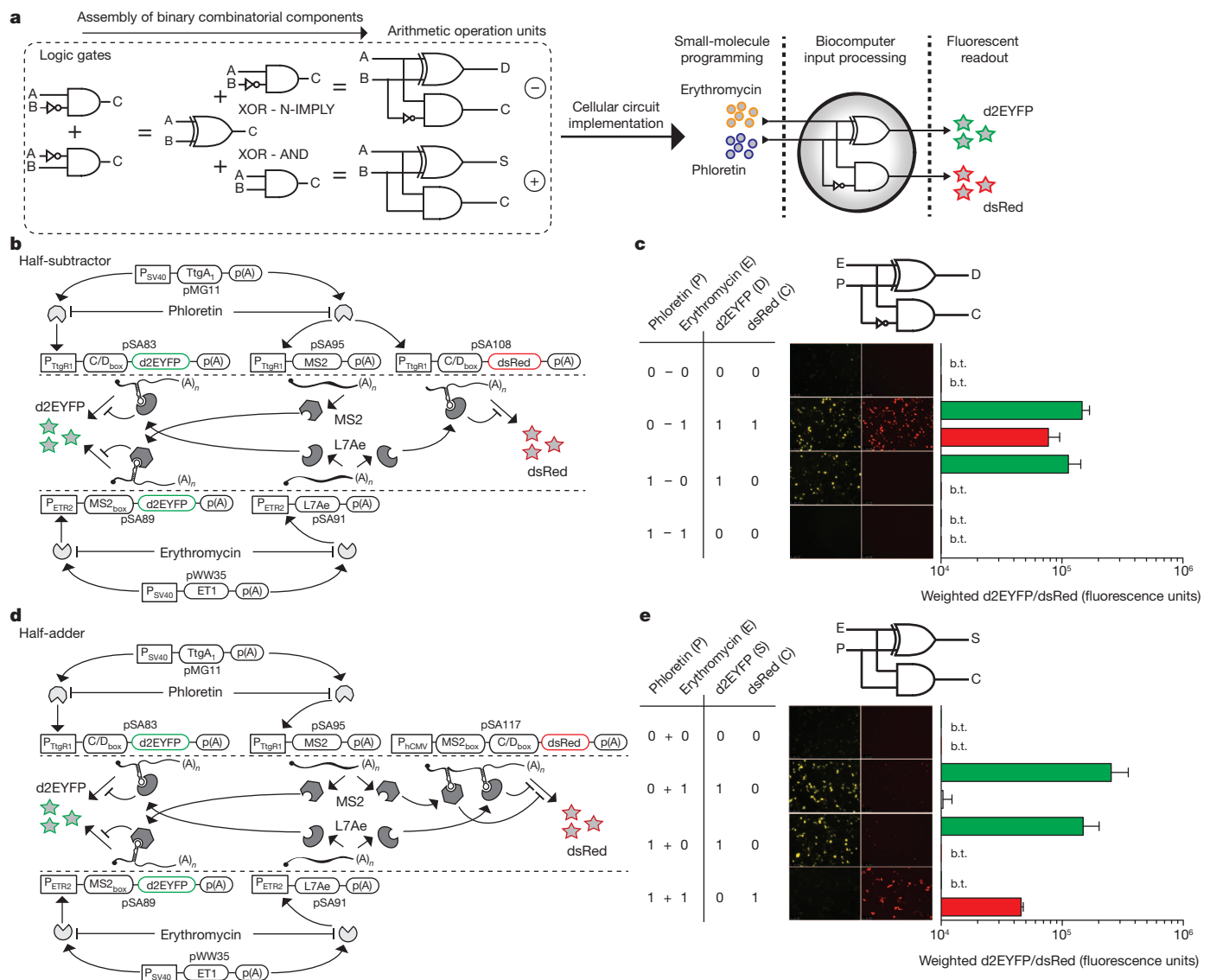
To test whether the transcription/translation-control devices could also be interconnected in a predictable manner to higher-order designer

Figure 3 | Design and computation characteristics of the synthetic mammalian XOR processor. **a**, Truth tables illustrating how the combination of two different N-IMPLY gates results in XOR processing, characterized by switching the output ON if exactly one of the input signals is present. **b**, Genetic switchboard of the XOR circuitry (see Supplementary Information for functional description). **c**, By combining the two input signals in accordance with the truth table, transfected HEK-293 cells are programmed to produce d2EYFP, following XOR computation logic: reporter gene expression is switched ON exclusively in the presence of either erythromycin or phloretin as shown by fluorescence microscopy and FACS analysis. Error bars represent s.d.; n = 3.

networks with complex computation logic we chose to engineer the XOR gate. XOR computations are particularly challenging to design because they integrate two different input signals and produce the output ON exclusively if one of the inputs is ON (Fig. 3a). Because of this complexity, XORs have so far only been constructed in non-cellular deoxyribozyme-based circuits²⁵ and as intercellular crosstalk in bacteria³ and yeast². Engineering HEK-293 cells with all components of both N-IMPLY gates (A ANDNOT B as well as B ANDNOT A) encoded on six different plasmids human cells produced d2EYFP with digital XOR expression logic after the addition of different combinations of the trigger-molecules erythromycin and phloretin (Fig. 3b, c). Although both d2EYFP expression units are completely transactivated by their respective transcription factors in the absence of input signals, the RNA controllers were able to inhibit translation efficiently and prevent d2EYFP expression (Fig. 3c). As required, the output of the XOR gate was also repressed in the presence of both input signals when both transcription factors were inactivated and also when neither of the two d2EYFP expression units was expressed. However, addition of either erythromycin or phloretin resulted in inactivation of the

corresponding transcription factor, whereas its counterpart remained active and induced its respective d2EYFP expression unit (Fig. 3c). For rigorous analysis of network dynamics and the contribution of individual system components to the computation behaviour we decomposed the XOR circuit by sequentially removing individual components; this resulted in distinct circuits that performed specific logic computations (Supplementary Fig. 4). Taking away either of the reporter genes resulted in N-IMPLY gates (Supplementary Fig. 4a, b), elimination of either transcription factor generated two distinct NOT gates (Supplementary Fig. 4c, d), and removal of either translation inhibitor produced two different gates with NAND logic (Supplementary Fig. 4e, f). These results underline the rational and predictable plug-and-play characteristic of the transcription/translation-control components and show that these individual components can be readily rewired into combinatorial structures to perform biocomputing activities ranging from simple NOT gates to complex XOR gates.

As a next step towards programmable single-cell biocomputers we added another layer of computation capacity to the XOR gate to design a half-subtractor, which is commonly used in digital electronics as a



binary arithmetic circuit capable of performing a subtraction of two bits. The half-subtractor was designed by combination of the XOR gate, which calculates the difference D, and an N-IMPLY gate to calculate the borrow C (Fig. 4a). The additional N-IMPLY gate (A ANDNOT B) was constructed by replacing the yellow fluorescent reporter gene of the aforementioned A ANDNOT B gate with dsRed so that its modified phloretin-regulated and L7Ae-controlled reporter unit would produce a red fluorescent output (Supplementary Fig. 5a). When placing this seven-component multi-processing device into human cells it performed precise two-input–two-output integration, showing half-subtractor calculation characteristics as predicted (Fig. 4b, c). Despite its high complexity, the half-subtractor performance was robust as indicated by the correlation of fluorescence microscopy and FACS analysis (Fig. 4c). In addition, in the presence of erythromycin, which programmed the half-subtractor to express both reporter genes, yellow and red fluorescence micrographs show almost perfect correlation, which suggests that the observed processing behaviour did indeed result from single-cell computations (Fig. 4c).

Besides the half-subtractor, the half-adder, which calculates the addition of two bits, is the second (bio)computer unit providing fundamental arithmetic operations. Together they manage any calculation in digital electronics. The design of a genetically encoded half-adder requires the combination of an XOR gate and an AND gate (Fig. 4a). The AND gate consists of a dsRed expression unit whose transcripts contain both MS2box and C/Dbox RNA motifs in their 5' UTR. They block the translation of dsRed on binding to MS2 or L7Ae, respectively. When MS2 transcription is set for repression by phloretin and L7Ae production is controlled in an erythromycin-repressible manner, pSA117-encoded dsRed is only expressed in the presence of both phloretin and erythromycin (Supplementary Fig. 5c). The pSA117-derived control vector lacking any RNA motifs showed constitutive dsRed expression (Supplementary Fig. 5d). Combination of this AND gate, which determines the carry C, with the aforementioned XOR gate (Fig. 3) calculating the sum S resulted in robust half-adder-specific computations in individual cells that can be programmed by a distinct set of input signals (Fig. 4d, e). Quantitative fluorescence profiling of entire circuit-transfected populations using a plate reader showed mean induction factors (OFF/ON) of 12 (d2EYFP) and 13 (dsRed) for the half-subtractor (Supplementary Figs 6a and 7a) and 14 (d2EYFP) and 4 (dsRed) for the half-adder (Supplementary Figs 6b and 7b). Furthermore, time-lapse fluorescence microscopy of entire circuit-transfected populations confirmed half-subtractor and half-adder processing dynamics in real time (Supplementary Movies 1 and 2).

In digital electronics the presence or absence of electrons represents the unique input signals, which allows multi-bit processing in a switchboard of multi-layered logic gates performing distinct calculations. The flow of electrons through the switchboard can be programmed to process desired algorithms. Although the design of electronic computers has achieved enormous processing power, living mammalian cells are in principle no less powerful, because they constantly process logic operations and perform countless physiological computations in parallel to coordinate endogenous metabolites and interface with environmental signals²⁶. Single-cell biocomputers are therefore compatible with any physiological trigger input that increases parallel processing power; they may be genetically programmed to execute encoded activities; they are scalable to tissue structures; and they are straightforward to interface with host metabolism to achieve therapeutic impact. Examples of synthetic biocomputing devices that sense signal input and process a coordinated therapeutic output such as sophisticated cancer kill switches integrating multi-input transformation signals^{10,11,27} were first tested in cell cultures, but pioneering one-input widgets such as T-cell population controllers²⁸, artificial insemination devices²⁹, blue-light-triggered glucose homeostasis for diabetes therapy¹³ and prosthetic gouty arthritis networks for the treatment of hyperuricemic disorders³⁰ have been used successfully

to perform therapeutic calculations when implanted and plugged into the metabolism of animals.

Programming single mammalian cells to perform multi-bit processing has been extremely challenging because the computations must operate in functional isolation from the cellular metabolism. Complex multi-bit processing devices have therefore been engineered in which individual processing components are encoded in single bacterial³ or yeast² cells, which communicate through chemical wires to perform computations in multicellular assemblies. We have successfully merged transcription and translation controllers in a combinatorial plug-and-play manner to achieve synthetic networks ranging from simple logic gates to complex XOR controllers as well as half-subtractors and half-adders executing fundamental arithmetic operations. Mammalian cells engineered with programmable genetic devices performing arithmetic calculations with similar precision, robustness and predictability to those of their digital electronics counterparts may enable the assembly of tissue-like biocomputers that could allow the design of complex human–machine interfaces and provide diagnostic information and therapeutic interventions in future gene-based and cell-based treatment strategies.

METHODS SUMMARY

Biocomputer design. Construction details for expression vectors are provided in Supplementary Information and Supplementary Table 1. HEK-293T cells (American Type Culture Collection, CRL-11268) were cultivated at 37 °C in DMEM medium (Invitrogen) supplemented with 10% FCS (lot no. PE01026P; Bioconcept) and 1% penicillin/streptomycin solution (Sigma-Aldrich) in a humidified atmosphere containing 5% CO₂. HEK-293 cells were (co)-transfected with biocomputer-encoding plasmids (Supplementary Table 2) by using a polyethylenimine-based protocol. At 6 h after transfection, the cells were re-seeded into 96-well plates containing erythromycin (2.7 µM; stock solution: 5 mg ml⁻¹ in 100% ethanol; Sigma Aldrich) and phloretin (50 µM; stock solution: 14.5 mM in 100% ethanol; Sigma Aldrich) as indicated and cultivated for a further 62 h before analysis (Supplementary Fig. 8).

Flow cytometry. Cell populations were analysed with a LSRII Fortessa flow cytometer (Becton Dickinson) equipped for d2EYFP (488-nm laser, 505-nm longpass filter and 530/30 emission filter (passband centred on 530 nm; passband width 30 nm)) and dsRed (561 nm laser, 505-nm longpass filter and 586/15 emission filter) detection and set to exclude dead cells and cell doublets. Each sample was spiked with AlignFlow alignment beads (A-7302; Life Technologies) as internal control, ensuring consistency of flow cytometry settings among different circuits and independent experiments (Supplementary Fig. 9). At least 10,000 cells were recorded per data set and analysed with FACSDiva (version no. 6.1.3; BD Biosciences). To score digital expression profiles of the biocomputer devices, transfected HEK-293 populations were gated for cells with high d2EYFP/yellow or dsRed/red expression/fluorescence levels beyond a threshold of 10⁴ arbitrary fluorescence units (see Supplementary Fig. 1 for control and FACS threshold-defining experiments and Supplementary Fig. 10 for raw data on circuit performance). The percentage of gated cells was multiplied by their median fluorescence, resulting in a weighted d2EYFP/dsRed expression value that correlated fluorescence intensity with cell number.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 4 November 2011; accepted 17 April 2012.

Published online 3 June 2012.

1. Khalil, A. S. & Collins, J. J. Synthetic biology: applications come of age. *Nature Rev. Genet.* **11**, 367–379 (2010).
2. Regot, S. *et al.* Distributed biological computation with multicellular engineered networks. *Nature* **469**, 207–211 (2011).
3. Tamsir, A., Tabor, J. J. & Voigt, C. A. Robust multicellular computing using genetically encoded NOR gates and chemical 'wires'. *Nature* **469**, 212–215 (2011).
4. Gitzinger, M., Kemmer, C., El-Baba, M. D., Weber, W. & Fussenegger, M. Controlling transgene expression in subcutaneous implants using a skin lotion containing the apple metabolite phloretin. *Proc. Natl Acad. Sci. USA* **106**, 10638–10643 (2009).
5. Weber, W. *et al.* Macrolide-based transgene control in mammalian cells and mice. *Nature Biotechnol.* **20**, 901–907 (2002).
6. Paraskeva, E., Atzberger, A. & Hentze, M. W. A translational repression assay procedure (TRAP) for RNA–protein interactions *in vivo*. *Proc. Natl Acad. Sci. USA* **95**, 951–956 (1998).
7. Saito, H., Fujita, Y., Kashida, S., Hayashi, K. & Inoue, T. Synthetic human cell fate regulation by protein-driven RNA switches. *Nature Commun.* **2**, 160 (2011).

8. Friedland, A. E. *et al.* Synthetic gene networks that count. *Science* **324**, 1199–1202 (2009).
9. Leisner, M., Bleris, L., Lohmueller, J., Xie, Z. & Benenson, Y. Rationally designed logic integration of regulatory signals in mammalian cells. *Nature Nanotechnol.* **5**, 666–670 (2010).
10. Xie, Z., Wroblewska, L., Prochazka, L., Weiss, R. & Benenson, Y. Multi-input RNAi-based logic circuit for identification of specific cancer cells. *Science* **333**, 1307–1311 (2011).
11. Culler, S. J., Hoff, K. G. & Smolke, C. D. Reprogramming cellular behavior with RNA controllers responsive to endogenous proteins. *Science* **330**, 1251–1255 (2010).
12. Weber, W. & Fussenegger, M. Molecular diversity—the toolbox for synthetic gene switches and networks. *Curr. Opin. Chem. Biol.* **15**, 414–420 (2011).
13. Ye, H., Daoud-El Baba, M., Peng, R. W. & Fussenegger, M. A synthetic optogenetic transcription device enhances blood-glucose homeostasis in mice. *Science* **332**, 1565–1568 (2011).
14. Kramer, B. P., Fischer, C. & Fussenegger, M. BioLogic gates enable logical transcription control in mammalian cells. *Biotechnol. Bioeng.* **87**, 478–484 (2004).
15. Win, M. N. & Smolke, C. D. Higher-order cellular information processing with synthetic RNA devices. *Science* **322**, 456–460 (2008).
16. Weber, W. *et al.* A synthetic time-delay circuit in mammalian cells and mice. *Proc. Natl Acad. Sci. USA* **104**, 2643–2648 (2007).
17. Weber, W., Daoud-El Baba, M. & Fussenegger, M. Synthetic ecosystems based on airborne inter- and intrakingdom communication. *Proc. Natl Acad. Sci. USA* **104**, 10435–10440 (2007).
18. You, L., Cox, R. S. III, Weiss, R. & Arnold, F. H. Programmed population control by cell–cell communication and regulated killing. *Nature* **428**, 868–871 (2004).
19. Basu, S., Gerchman, Y., Collins, C. H., Arnold, F. H. & Weiss, R. A synthetic multicellular system for programmed pattern formation. *Nature* **434**, 1130–1134 (2005).
20. Greber, D. & Fussenegger, M. An engineered mammalian band-pass network. *Nucleic Acids Res.* **38**, e174 (2010).
21. Danino, T., Mondragon-Palomino, O., Tsimring, L. & Hasty, J. A synchronized quorum of genetic clocks. *Nature* **463**, 326–330 (2010).
22. Tigges, M., Marquez-Lago, T. T., Stelling, J. & Fussenegger, M. A tunable synthetic mammalian oscillator. *Nature* **457**, 309–312 (2009).
23. Ellington, A. D. & Szostak, J. W. *In vitro* selection of RNA molecules that bind specific ligands. *Nature* **346**, 818–822 (1990).
24. Saito, H. *et al.* Synthetic translational regulation by an L7Ae-kink-turn RNP switch. *Nature Chem. Biol.* **6**, 71–78 (2010).
25. Stojanovic, M. N. & Stefanovic, D. Deoxyribozyme-based half-adder. *J. Am. Chem. Soc.* **125**, 6673–6676 (2003).
26. Bandyopadhyay, S. *et al.* Rewiring of genetic networks in response to DNA damage. *Science* **330**, 1385–1389 (2010).
27. Nissim, L. & Bar-Ziv, R. H. A tunable dual-promoter integrator for targeting of cancer cells. *Mol. Syst. Biol.* **6**, 444 (2010).
28. Chen, Y. Y., Jensen, M. C. & Smolke, C. D. Genetic control of mammalian T-cell proliferation with synthetic RNA regulatory systems. *Proc. Natl Acad. Sci. USA* **107**, 8531–8536 (2010).
29. Kemmer, C. *et al.* A designer network coordinating bovine artificial insemination by ovulation-triggered release of implanted sperms. *J. Control. Release* **150**, 23–29 (2011).
30. Kemmer, C. *et al.* Self-sufficient control of urate homeostasis in mice by a synthetic circuit. *Nature Biotechnol.* **28**, 355–360 (2010).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank R. Singer for providing pMS2dIFG, M. Tigges for generous advice, E. Gutzwiller for experimental support, and M. Dessing and V. Jäggin for assistance with flow cytometry. This work was supported by the Swiss National Science Foundation (grant no. 31003A-126022) and in part by EC Framework 7 (Persist).

Author Contributions S.A., D.A., M.M., M.W. and M.F. designed the project, analysed results and wrote the manuscript. S.A., D.A. and M.M. performed the experimental work.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to M.F. (fussenegger@bsse.ethz.ch).

METHODS

Biocomputer components. Comprehensive design and construction details for all expression vectors are provided in Supplementary Table 1. Key plasmids include the following. pWW35 (ref. 5) encodes constitutive expression of the macrolide-dependent transactivator (ET1; P_{SV40} -ET1-pA). pMG11 (ref. 4) harbours a constitutive expression unit encoding the phloretin-dependent transactivator (TtgA₁; P_{SV40} -TtgA₁-pA). pSA83 encodes a phloretin-responsive TtgA₁-triggered P_{TigR1} -driven d2EYFP expression unit harbouring a C/Dbox in the 5' UTR that is targeted by the translation suppressor protein L7Ae²⁴ (P_{TigR1} -C/D_{box}-d2EYFP-pA; GenBank ID JQ624673). pSA89 encodes an erythromycin-responsive ET1-triggered P_{ETR2} -driven d2EYFP expression unit harbouring a MS2box in the 5' UTR that is targeted by the translation suppressor protein MS2 (ref. 6) (P_{ETR2} -MS2_{box}-d2EYFP-pA; GenBank ID JQ624674). pSA91 encodes an erythromycin-responsive ET1-triggered P_{ETR2} -driven L7Ae expression unit (P_{ETR2} -L7Ae-pA; GenBank ID JQ624675). pSA95 encodes a phloretin-responsive TtgA₁-triggered MS2 expression unit (P_{TigR1} -MS2-pA; GenBank ID JQ624676). pSA108 encodes a phloretin-responsive TtgA₁-triggered P_{TigR1} -driven dsRed expression unit harbouring a C/Dbox in the 5' UTR that is targeted by the translation suppressor protein L7Ae²⁴ (P_{TigR1} -C/D_{box}-dsRed-pA). pSA117 encodes a constitutive P_{hCMV} -driven dsRed expression unit harbouring both MS2box and C/Dbox RNA motifs in its 5' UTR that are targeted by the translation suppressor proteins MS2 (ref. 6) and L7Ae²⁴, respectively (P_{hCMV} -MS2_{box}-C/D_{box}-dsRed-pA). The plasmid composition of each circuit is provided in Supplementary Table 2.

Cell culture, transfection and gene regulation. HEK-293T cells (American Type Culture Collection, CRL-11268) were cultivated at 37 °C in DMEM medium (Invitrogen) supplemented with 10% FCS (lot no. PE01026P; Bioconcept) and 1% penicillin/streptomycin solution (Sigma-Aldrich) in a humidified atmosphere containing 5% CO₂. HEK-293 cells were (co)-transfected using an optimized polyethylenimine-based protocol. In brief, a transfection solution containing 4 µg of plasmid DNA mixtures, 800 µl of FCS-free DMEM and 12 µl of polyethylenimine (1 mg ml⁻¹ in water; Polysciences) was incubated for 30 min at 22 °C before it was added dropwise to 4 × 10⁵ HEK-293 cells seeded in each well of a six-well plate 14 h before transfection (see Supplementary Table 2 for a detailed composition of plasmid DNA mixtures). At 6 h after transfection, the cells were detached for 1 min using 200 µl of trypsin/EDTA (Biowest), washed once in 1 ml

of FCS-containing DMEM, resuspended in 0.75 ml of FCS-containing DMEM, re-seeded into 12 wells of a 96-well plate containing the trigger molecules erythromycin (2.7 µM; stock solution: 5 mg ml⁻¹ in 100% ethanol; Sigma Aldrich) and phloretin (50 µM; stock solution: 14.5 mM in 100% ethanol; Sigma Aldrich) as indicated and cultivated for a further 62 h before analysis (Supplementary Fig. 8).

Fluorescence imaging. Fluorescence and time-lapse microscopy was performed with an inverted fluorescence microscope (DMI 6000B; Leica Microsystems) equipped with an incubation chamber, a DFC350FX R2 digital camera (Leica), a 10× objective (objective HC PL FL 10×/0.30 PH1 -/D 11.0; Leica), a 495/535-nm (d2EYFP) and 580/630-nm (dsRed) excitation/emission filter set and LAS AF imaging software (FW4000-TZ; Leica). Identical settings including exposure times of 1 s for d2EYFP and dsRed were used for all fluorescence micrographs.

Flow cytometry. All engineered cell populations were analysed with a Becton Dickinson LSRII Fortessa flow cytometer equipped for d2EYFP (488 nm laser, 505 nm longpass filter, 530/30 emission filter (passband centred on 530 nm; passband width 30 nm)) and dsRed (561 nm laser, 505 longpass filter and 586/15 emission filter) detection and set to exclude dead cells and cell doublets. Before analysis, each sample was spiked with AlignFlow alignment beads (A-7302; Life Technologies), serving as an internal control and ensuring consistency of flow cytometry settings between different circuits and independent experiments (Supplementary Fig. 9). At least 10,000 cells were recorded in each data set and analysed with BD FACSDiva (version no. 6.1.3; BD Biosciences). To score digital expression profiles of the biocomputer devices, transfected HEK-293 cell populations were gated for cells with high d2EYFP/yellow or dsRed/red expression/fluorescence levels beyond a threshold of 10⁴ arbitrary fluorescence units (see Supplementary Fig. 1 for control and FACS threshold-defining experiments and Supplementary Fig. 10 for raw data on circuit performance). The percentage of gated cells was multiplied by their median fluorescence, resulting in a weighted d2EYFP/dsRed expression value that correlated fluorescence intensity with cell number.

Plate-reader-based fluorescence analysis. Fluorescence of entire circuit-transfected cell populations was profiled with a monochromatic Tecan Infinite 200pro plate reader with excitation and emission set to 488/9 nm and 535/20 nm for d2EYFP, and 560/9 nm and 595/20 nm for dsRed, respectively. Autofluorescence of mock-transfected cells was deducted from circuit-derived fluorescence.

CORRIGENDUM

doi:10.1038/nature11168

Corrigendum: Fetal load and the evolution of lumbar lordosis in bipedal hominins

Katherine K. Whitcome, Liza J. Shapiro & Daniel E. Lieberman

Nature **450**, 1075–1078 (2007); doi:10.1038/nature06342

In this Letter, Fig. 4a appeared incorrectly (see corrected figure below). The correct equation used to calculate vertebral wedging is: $2\arctan\{[(\text{centrum dorsal height} - \text{centrum ventral height})/2]/\text{centrum anteroposterior diameter}\}$, but the correction constant of 2 was inadvertently omitted. As a result, the australopithecine wedging angles reported are half the correct values. However, after correction, our original inference of similarity in sex-related patterns of vertebral wedging among bipedal hominins remains strongly supported. We thank K. Ostrofsky, S. A. Williams and S. Churchill for drawing this error to our attention. Figure 4a has been corrected in the HTML and PDF versions online.

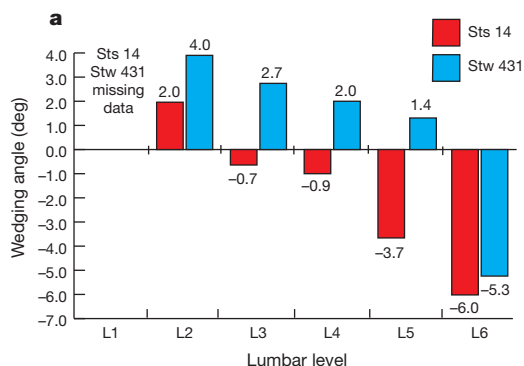


Figure 1 | This is the corrected Fig. 4a of the original Letter.

RESEARCH HIGHLIGHTS

Selections from the
scientific literature

GENE THERAPY

Quit smoking with a shot

Injecting a gene that encodes an anti-nicotine antibody into mice blocks the addictive chemical from entering the brain.

Ronald Crystal at Weill Cornell Medical College in New York and his team injected mice with a vaccine containing the gene. Viruses in the vaccine shuttled the gene to the liver. Once there, liver cells generated the antibodies, which bind to nicotine in the blood and prevent it from crossing the blood–brain barrier.

Mice that were administered the vaccine and then nicotine had only 15% as much brain nicotine as unvaccinated animals. Moreover, vaccinated mice did not show the increased blood pressure and heart rate that unvaccinated mice experienced after receiving nicotine.

Sci. Transl. Med. 4, 140ra87 (2012)

NEUROSCIENCE

fMRI translates thoughts to words

With minimal training, people can communicate silently through a functional magnetic resonance imaging (fMRI) machine by performing mental-imagery tasks that correspond to the 26 letters of the English alphabet.

Bettina Sorger at Maastricht University in the Netherlands and her team asked six healthy volunteers to use a set of three mental tasks. Each task could be performed for three different durations and

its onset delayed by three different lengths of time. This generated 27 different fMRI activation patterns, one for each letter and one for a space. In a one-hour fMRI scanning session, the volunteers spelled out their answers to open questions (**examples pictured**) by carrying out these mental tasks, with the help of a visual letter-encoding guide. A decoding algorithm translated the fMRI signals back into letters (**pictured, top three rows**) in real-time.

The technique could be used by patients who cannot

communicate physically but are still conscious.

Curr. Biol. <http://dx.doi.org/10.1016/j.cub.2012.05.022> (2012)

ENVIRONMENTAL SCIENCE

Future ozone from planes and boats

Emissions from aircraft and shipping are anticipated to be the dominant sources of increases in ground-level ozone, which can trigger respiratory disease.

Didier Hauglustaine and

seal (pictured).

In species that compete for a harem, greater body mass — useful in battle — is associated with reduced penis length and testis mass. This indicates that males that invest more in pre-copulatory strategies such as increased body size may do so at the expense of post-copulatory ones such as improving the quality of their ejaculates.

Evolution <http://dx.doi.org/10.1111/j.1558-5646.2012.01713.x> (2012)



EVOLUTION

Fighters invest in body, not testes

Males that fight for the right to mate seem to invest less in testis size and more in body size compared with animals that do not fight — highlighting the trade-off that occurs between different strategies for reproductive success.

John Fitzpatrick at the University of Western Australia in Perth and his colleagues collated data on penis length, testis mass and the body mass of males relative to females in sea lions, walrus and seals, such as the Antarctic fur

Brigitte Koffi of the Laboratory of Climate and Environmental Sciences in Gif-sur-Yvette, France, used a global climate model to calculate the relative contributions of road transport, aircraft and shipping to the projected increase by 2050 in ozone pollution in Europe and the United States.

Nitrogen oxide emissions from aircraft will contribute more than 30% of the transport-induced increase in peak summertime surface ozone. Shipping emissions will, in some regions, account for up to 60% of the increase. Thanks to cars becoming cleaner, road transport will probably contribute little.

Geophys. Res. Lett. <http://dx.doi.org/10.1029/2012GL052008> (2012)

"Where did you spend your most recent vacation?"

- - U D - D E S T -
I A V C A P C U U I
A B C F B Y D R V A
- B U D A P E S T -

"What did you like most in BUDAPEST?"

- S W N - E D F U E -
A U X L A G X E V D A
J T Y M D F M G S C R
- S Y N A G O G U E -

NEURODEGENERATION

Silencing Huntington's gene

A molecule that lowers the expression of the mutant protein that causes Huntington's disease reverses symptoms of the disorder in mouse models of the disease.

Don Cleveland at the University of California, San Diego, and his team infused a short string of nucleotides that blocks the expression of the huntingtin gene into the brains of mice bearing the disease-causing mutation. The molecule persisted in brain tissue for more than three months, and levels of mutant protein were reduced for up to four months after the infusion. Nine months post-treatment, the mice continued to show decreased anxiety and improved motor function.

When the oligonucleotide was infused into the cerebrospinal fluid of Rhesus macaques, levels of huntingtin messenger RNA fell in multiple brain regions. *Neuron* 74, 1031–1044 (2012)

BIOPHYSICS

How cells maintain chains

Long protein chains in the form of microtubules and actin filaments give cells their shape and rigidity, and also perform chores such as yanking apart copies of chromosomes. Two mathematical models explain how the length of these chains is kept in check by molecular motor-proteins that 'walk' along the chains.

Karsten Kruse and his team at Saarland University in Saarbrücken, Germany, modelled the growth and shrinkage of actin filaments, which occurs at both ends of the chain. Erwin Frey and his group at the Ludwig Maximilian University in Munich, Germany, focused on microtubules, which are trimmed by molecular motors at a single tip.

The teams found that

molecular motors control chain length by forming 'traffic jams' at specific places along the chain. At these spots, protein building blocks are continuously gained and shed so that the overall length remains stable. The location of the traffic jam determines the length of the chain, which terminates at the jam.

Phys. Rev. Lett. 108, 258103; 258104 (2012)

NEUROSCIENCE

DNA methylation controls memory

As animals age, the addition of methyl groups to DNA — a mechanism that controls gene expression — decreases in the brain. This reduction causes cognitive decline, researchers now show.

Hilmar Bading and his colleagues at the University of Heidelberg in Germany showed that a gene central to DNA methylation, *Dnmt3a2*, controls cognitive abilities in mice. When the researchers overexpressed *Dnmt3a2* in the hippocampal brain area of elderly mice, the animals regained their cognitive skills in two long-term-memory tests. Reducing hippocampal expression of the gene caused young adult mice to perform poorly in the tests compared with controls of the same age. *Nature Neurosci.* <http://dx.doi.org/10.1038/nn.3151> (2012)

CLIMATE CHANGE

Lopsided warming north to south

The Northern Hemisphere is warming more quickly than the Southern Hemisphere thanks to the uneven distribution of land and sea around the globe, a study confirms.

Land masses warm more quickly than the ocean — in part because more heat is required to raise the temperature of water than that of land, and in part owing to the cooling effects of sea surface evaporation. Yangyang Xu and



COMMUNITY CHOICE

The most viewed papers in science

ECOLOGY

Illuminating invertebrate habitats

HIGHLY READ
on rsbl.royalsociety-publishing.org
in May

Street lamps can shape the surrounding invertebrate community, attracting predators and scavengers day and night.

Thomas Davies and his team at the University of Exeter in Penryn, UK, set traps under and between 14 lamp posts on a street in a town in Cornwall, and collected samples before each sunrise and sunset for three days. The researchers classified the invertebrates on the basis of their feeding strategies: predators, scavengers, grazers, parasites and detritivores, which consume rotting material. Overall, there were more invertebrates under the lamps than between them, but predators and scavengers in particular were more abundant. Harvestmen, ants and ground beetles preferred the grass under the lamp posts both day and night, indicating that they were not simply drawn by the light.

Light pollution could affect the structure and function of ecosystems, the researchers suggest.

Biol. Lett. <http://dx.doi.org/10.1098/rsbl.2012.0216> (2012)



Veerabhadran Ramanathan of the Scripps Institution of Oceanography in La Jolla, California, used climate simulations to confirm that these effects can explain the global pattern of temperature change seen during periods of warming — in 1910–40 and 1975–2005 — as well as during a cooling period in 1940–75. The climate-model runs predict drastic changes in rainfall, including intense droughts in Africa, Western North America, the Amazon and Australia, and a boost to the Indian monsoon. *Geophys. Res. Lett.* <http://dx.doi.org/10.1029/2012GL052116> (2012)

NEUROSCIENCE

Gene blocks effects of stress

Chronic stress can contribute to depression, but its effects can be blocked in rats by upregulating a single gene in one brain region.

A gene called neuritin encodes a protein that is involved in the growth of

neuronal branches. Hyeon Son and Ronald Duman at Yale University in New Haven, Connecticut, and their team studied the gene at work in a rat model of chronic stress. The researchers found that exposure to stress decreased neuritin expression in a region of the brain called the hippocampus, and that the effect was reversed by an antidepressant. Boosting neuritin protein levels increased the amount of neuronal branching and prevented stress from affecting the animals' behaviour. Silencing the neuritin gene brought on depression-like behaviours.

Chronic stress leads to atrophy of hippocampal neurons, which can increase susceptibility to mood disorders, the authors suggest. *Proc. Natl Acad. Sci. USA* <http://dx.doi.org/10.1073/pnas.1201191109> (2012)

➔ **NATURE.COM**

For the latest research published by Nature visit:
www.nature.com/latestresearch